

February 14, 2003 13:33 27

(without alignments)
8796.662 Million seconds

00-395B-1
45

ENTITY_NUC

1362 seqs, 153338301

100

000000002 : 1171

Minimum Match 0%

string first 45

sued_Patents NA.†

```
/cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

by analysis of the total score distribution.

Results

Description					ID	DB	Length	Match	C
Sequence 1, Appl1	US-09-345-882-1	4	162450	6.7	150	C 1			
Sequence 10, App	US-09-371-211-10	4	98844	6.7	149.6	C 2			
Sequence 12, Appl	US-09-731-319C-12	2	4129	6.6	148.8	C 3			
Sequence 12, App	US-09-224-834-12	4	6.5	144.8	C 4				
Sequence 11, Appl	US-09-750-580-1	4	81001	6.4	144.8	C 5			
Sequence 11, Appl	US-08-687-080-111	2	2448	6.4	143.8	C 6			
Sequence 1, Appl1	US-09-210-748A-3	4	72928	6.4	143.8	C 7			
Sequence 1, Appl1	US-09-009-913-1	4	99500	6.4	143	C 8			
Sequence 10, Appl	US-09-728-096-10	4	29629	6.3	142.6	C 9			
Sequence 7, Appl1	US-08-941-587-7	4	45716	6.3	142.2	C 10			
Sequence 5, Appl1	US-08-965-048-5	4	45989	6.3	142.2	C 11			
Sequence 6, Appl1	US-08-323-434B-1	1	31571	6.3	141.2	C 12			
Sequence 1, Appl1	US-08-658-136-1	3	53577	6.3	141.2	C 13			
Sequence 2, Appl1	US-08-975-080-35	4	14796	6.3	140.8	C 14			
Sequence 35, Appl1	US-09-630-706-10	4	14796	6.3	140.8	C 15			
Sequence 3, Appl1	US-09-496-694B-3	4	14796	6.3	140.8	C 16			
Sequence 3, Appl1	US-09-817-180-3	4	45546	6.2	139.4	C 17			
Sequence 6, Appl1	US-09-385-986-332	4	84495	6.2	139.4	C 18			
Sequence 332, App	US-09-741-154-3	4	15389	6.2	138.4	C 19			
Sequence 3, Appl1	US-09-813-917-3	4	59065	6.2	138.4	C 20			
Sequence 3, Appl1	US-09-978-197-3	4	9365	6.2	138.2	C 21			
Sequence 8, Appl1	US-09-608-285A-8	4				C 22			
						C 23			
						C 24			
						C 25			
						C 26			
						C 27			

C	28	138.2	6.2	9355	4	US-09-350-836B-8	Sequence 8, Appl
C	29	138.2	6.2	9365	4	US-09-370-265-8	Sequence 8, Appl
C	30	138.2	6.2	14747	4	US-09-608-285A-2	Sequence 8, Appl
C	31	138.2	6.2	15977	4	US-09-608-285A-52	Sequence 42, Appl
C	32	136.8	6.1	9734	4	US-09-347-114A-80	Sequence 59, Appl
C	33	136.8	6.1	49136	4	US-09-422-869-1	Sequence 80, Appl
C	34	136.6	6.1	55990	4	US-09-045-159-129	Sequence 1, Appl
C	35	136.6	6.1	162450	4	US-09-345-882-1	Sequence 129, Appl
C	36	136.4	6.1	11558	5	PCT-US93-06251-23	Sequence 1, Appl
C	37	136.2	6.1	1701	4	US-09-078-294-9	Sequence 23, Appl
C	38	136.2	6.1	17041	1	US-08-076-011-1	Sequence 9, Appl
C	39	136.2	6.1	87350	3	US-08-781-891-79	Sequence 1, Appl
C	40	135.8	6.1	87543	4	US-09-791-211-3	Sequence 79, Appl
C	41	135.8	6.0	36159	4	US-09-749-886-3	Sequence 3, Appl
C	42	135.6	6.0	14581	4	US-08-520-8730-4	Sequence 3, Appl
C	43	135.6	6.0	22481	4	US-08-367-841A-43	Sequence 43, Appl
C	44	135.6	6.0	22481	5	PCT-US95-07201-43	Sequence 43, Appl
C	45	135.6	6.0	22484	4	US-09-875-223-2	Sequence 43, Appl

RESULT 1

```

US-09-345-882-1/c
/ Sequence 1, Application US/09345882
/ Patent No. 639373X
/ GENERAL INFORMATION:
/ APPLICANT: Bougueleret, Lydie
/ TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP
/ FILE REFERENCE: GENSET.031A AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
/ CURRENT APPLICATION NUMBER: US/09/345,882
/ PRIOR FILING DATE: 1999-06-30
/ PRIOR FILING DATE: 1998-06-30
/ PRIOR APPLICATION NUMBER: US 60/091,315
/ PRIOR FILING DATE: 1998-12-10
/ NUMBER OF SEQ ID NOS: 140
/ SOFTWARE: Patent.pm
/ SEQ ID NO 1
/ LENGTH: 162450
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 72794
/ OTHER INFORMATION: 5-124-273 : polymorphic base A or G
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 88073
/ OTHER INFORMATION: 5-127-261 : polymorphic base A or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 90842
/ OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 93714
/ OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 97122
/ OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 97152
/ OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 99098
/ OTHER INFORMATION: 5-130-257 : polymorphic base A or G
/ FEATURE:

```

NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108477
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele

LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177

FEATURE:

	Query Match	6.43;	Score 144.8;	DB 4:	Length 81001;
	Best Local Similarity	83.78;	Pred. No. 1,9e+20;		
	Matches 164; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0;
OY	1996 CTACTAAATAATATTAATAATTAGCTGGGTGTTGGTGGTGCCCTGAATACCCAGCTACAT				
Db	64921 CTAATTAATAATNCAAAAATTTAGCCGGCATGTGGCAGAGGCGCTAATATCCAGCTACTGC				
OY	2056 GGGAGCGCTGGGCGATAGAGAATCACTTGAACCTGGGGGGTGGGAGTTCAGTAGAGCTGAGA				
Db	64861 AGGAGGCTAGAGCAGAGGATCATCTTGAACCCGGGAGGGGAGGTTCAGTAGAGCTGAGA				
OY	2116 TTGGCGCACTAGCACMTCMACTCTGTCTGGCACAAGAGAGAACCTCCGCTCTAAAAACAACA				
Db	64801 TTGGCCCACTGACACTCCAGCCTGAGGGGACAAGACGAGAGCTCTGTATCAAAAAAAAAAAC				
OY	2176 AAAAAACCACTAACCA 2191				
Db	64741 AAAAAATTAATAATA 64726				

RESULT 6
US-08-687-080-111
; Sequence 111, Application US/08687080
; Patent No. 5965427X h.u.v.w.
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

```

? PRIOR APPLICATION NUMBER: 60/069,416
? PRIOR FILING DATE: 1997-12-18
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: fastseq for Windows Version 3.0
? SEQ ID NO 3
? LENGTH: 7680
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-210-748A-3

Query Match          6.4%; Score 143.8; DB 4; Length 7680;
Best Local Similarity 74.5%; Pred. No. 2e+20; Indels 0; Gaps 0.
Matches 181; Conservative 0; Mismatches 62;

QY   2001 AAAAAATTGAATTTAGCTGCGTGTGGTGTTGGTGGTGCTGTAAATCCACGACTACATGGAG 2060
      |||||
DB   5556 AAAAATCAAAAATNTACCTGGTGGTGTGGTGCACCCTTGAATCCACGACTACGCGGAG 5615
      |||||

QY   2061 CCTGAGCATGAGAAATTCACCTTGAACTCGGAGGTGGAGGTTTCAGTAGCCTGACATTTGCC 2120
      |||||
DB   5616 GCTGGGAGAGAGAAATACCTTGAACCCAGSAGGTGAGCGTTGCACGTGAGCTGAGATTGTG 5675
      |||||

QY   2121 CCACGACACTCCAATCTGTGTGCGCAGAAGTGAACCTCCGTCACAAAACATAAAACAAAAA 2180
      |||||
DB   5676 CCACACTACACTCCAGCGCTGGGTGACAAGAGCGAATCTCCGTCACAAAAAAGAAATTA 5735
      |||||

QY   2181 ACCACATACATAAATTTATCATCTGCAACCACTTTTACGTGCGGCAATTCACATCTCA 2240
      |||||
DB   5736 TATATATATATAGTTCACATCATCTCACTGTGCTTTCATTAATCTGCTGTAAATCCTCACA 5795
      |||||

QY   2241 TGT 2243
      ||
DB   5796 AGT 5798

```


Db 13557 AAGAAAAAAGAAATTAATAT 13579

RESULT 12

US-08-965-048-5/c Application US/08965048

Sequence 5, Patent No. 6323244

GENERAL INFORMATION:

APPLICANT: Chen, Hong

APPLICANT: Freimer, Nelson

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC DISORDERS

FILE REFERENCE: 7853-093

CURRENT APPLICATION NUMBER: US/08/965,048

CURRENT FILING DATE: 1997-11-05

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 45716

TYPE: DNA

ORGANISM: Homo sapiens

US-08-965-048-5

Query Match
Best Local Similarity 6.3%; Score 142.2; DB 4; Length 45716;
Matches 162; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1996 CTACTAAAAATATATAATTTGCTGGTGTGTGGTGGTCCCTGTAATCCAGCTACAT 2055

Db 29605 CTATTAAAAATACAAAAATTTAGCCAGGTGTGGTGGTGTCTGTAATCCAGCTGCTA 29546

QY 2056 GGGAGCCTGAGGCGATGAGATCACTTGTGAACTCGGAGAGTGGAGGTTGCTGAGCTGAGA 2115

Db 29545 GGGAGCCTGAGGCGATGAGATCACTTGTGAACTCGGAGAGTGGAGGTTGCTGAGCTGAGA 29486

QY 2116 TTGGCCCACTGACCTCCAGTCTGTGCGCAAGAGTGAAGTCCGCTCAAAAAACAACA 2175

Db 29485 TTGGCCCACTGACCTCCAGTCTGTGCGCAAGAGTGAAGTCCGCTCAAAAAACAACA 29426

QY 2176 AAAAAACCATATAC 2190

Db 29425 AAGAGAGAAAAAGAC 29411

RESULT 13

US-08-965-048-6/c

Sequence 6, Application US/08965048

Patent No. 6323244

GENERAL INFORMATION:

APPLICANT: Chen, Hong

APPLICANT: Freimer, Nelson

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC DISORDERS

FILE REFERENCE: 7853-093

CURRENT APPLICATION NUMBER: US/08/965,048

CURRENT FILING DATE: 1997-11-05

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6

LENGTH: 45989

TYPE: DNA

ORGANISM: Homo sapiens

US-08-965-048-6

Query Match
Best Local Similarity 6.3%; Score 142.2; DB 4; Length 45989;
Matches 162; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1996 CTACTAAAAATATATAATTTGCTGGTGTGTGGTGGTCCCTGTAATCCAGCTACAT 2055

Db 29719 CTATTAAAAATACAAAAATTTAGCCAGGTGTGGTGGTGTCTGTAATCCAGCTGCTA 29660

QY 2056 GGGAGCCTGAGGCGATGAGATCACTTGTGAACTCGGAGAGTGGAGGTTGCTGAGCTGAGA 2115

Db 29659 GGGAGCCTGAGGCGATGAGATCACTTGTGAACTCGGAGAGTGGAGGTTGCTGAGCTGAGCCGAGA 29600

QY 2116 TTGGCCCACTGACCTCCAGTCTGTGCGCAAGAGTGAAGTCCGCTCAAAAAACAACA 2175

Db 29599 TTGGCCCACTGACCTCCAGTCTGTGCGCAAGAGTGAAGTCCGCTCAAAAAACAACA 29540

QY 2176 AAAAAACCATATAC 2190

Db 29539 AAGAGAGAAAAAGAC 29525

RESULT 14

US-08-323-443B-1/c

Sequence 1, Application US/08323443B

Patent No. 5654170

GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W.

APPLICANT: LANDES, GREGORY M.

APPLICANT: BURN, TIMOTHY C.

APPLICANT: CONNORS, TIMOTHY D.

APPLICANT: DACKOWSKI, WILLIAM R.

APPLICANT: GERMINO, GREGORY

APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dady & Dady PC

STREET: 805 Third Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/323,443B

FILING DATE: 12-OCT-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, S. Peter

REGISTRATION NUMBER: 25,351

REFERENCE/DOCKET NUMBER: 0372/0A462

TELEPHONE: (212) 527-7700

TELEFAX: (212) 753-6237

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 31571 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: PKD1 GENOMIC

US-08-323-443B-1

Query Match
Best Local Similarity 6.3%; Score 141.2; DB 1; Length 31571;
Matches 158; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1996 CTACTAAAAATATATAATTTGCTGGTGTGTGGTGGTCCCTGTAATCCAGCTACAT 2055

Db 1062 CTACTAAAAATATATAATTTGCTGGTGTGTGGTGGTCCCTGTAATCCAGCTACAT 1003

QY 2056 GGGAGCCTGAGGCGATGAGATCACTTGTGAACTCGGAGAGTGGAGGTTGCTGAGCTGAGA 2115

QY 1075 CACCTGGCCCTGCTCCCTCCGCTCCGAGATCCCACTGGACCTGG 1125
 DB 492 CACCTGGCTGCTGCTCCCACTGCTGAGATACCACTGGCTCTGG 542
 RESULT 3
 A1510675 588 bp mRNA linear EST 12-MAR-1999
 LOCUS A1510675
 DEFINITION vx91b07.y1 Soares_thymus_2NBMT Mus musculus cDNA clone
 IMAGE:1282549 5', mRNA sequence.
 ACCESSION A1510675
 VERSION A1510675.1 GI:4409580
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:674349
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 398.
 Location/Qualifiers
 1..588
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="1282549"
 /clone_lib="Soares_thymus_2NBMT"
 /sex="male"
 /tissue_type="Thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATGTCAGAGGAGGCGCGCTTTTCTTTTCTTTTCTTTTCTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M. Fatima Bernaldo."
 BASE COUNT 126 a 165 c 165 g 130 t 2 others
 ORIGIN
 Query Match 10.6%; Score 237.4; DB 9; Length 588;
 Best Local Similarity 71.9%; Pred. No. 6,7e-30;
 Matches 353; Conservative 0; Mismatches 128; Indels 10; Gaps 3;

QY 838 ---GGCAGTGGGAAGAACAAGGCCGAGCAGCAGTGGCCCGAAGCCTGTGTCAGACC 894
 DB 284 AGTGGCAATTTTGAAGAACAGAGCCCGAGAGTGAAGCCTTAAGCCAGTGTCCAGACC 343
 QY 895 AAGGAGAGCCAGAGGAGCTGCGCCCGGAGAGTGAAGCTGAGCTGGGCGCAGCAGGAGC 954
 DB 344 AAGGAGAGCCAGAGTCACTATACCTAGATGAGCAGAAAGTGGCGCAGCAGTGTGGG---- 399
 QY 955 GTTCCGCCCCCTGCTGGCCCTCCCAAGTACCCAGCTCCACGAAAGTGAAGCAGAG 1014
 DB 400 --TCTTCCCTTCATCCCTTCCAGTGAAGCCCAAGCTTAACGAAAGAGGATGAGTG 457
 QY 1015 TGTGCGCTGTGTGGAGAGCGCGGAGACTATCTGCTGTACGCGCTG--CCCTGGGCGCTT 1073
 DB 458 TGGCGTGTGACAGACCGAGGAGTGAAGCTTATCTGTGTGACGCGCTGTCCCGGCGCTT 517
 QY 1074 CCACCTGGCCCTGCTGCTCCCTCCGCTCCGAGAGATCCCACTGGAGCCTGGAGCTGCTC 1133
 DB 518 CACCTGCTGCTGCTGCTCCCACTCTGCAAGAGACCCAGTGGCTTTTGAAGATTTGCT 577
 QY 1134 CAGCTGCTGCTGCT 1144
 DB 578 CTGCTTGTGCTTC 588
 RESULT 4
 HSDML04M6 200 bp DNA linear GSS 17-AUG-1998
 LOCUS HSDML04M6
 DEFINITION Homo sapiens DNA for trapped exon (ID MDL04M6), genomic survey
 sequence.
 ACCESSION AJ001114
 VERSION AJ001114.1 GI:3426119
 KEYWORDS GSS.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 200)
 AUTHORS Scott,H.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-AUG-1997) Scott H.S., Department of Genetics and
 Microbiology, University of Geneva Medical School, 1 rue Michel
 Servet, Geneva, 1211, SWITZERLAND
 2 (bases 1 to 200)
 REFERENCE Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S.,
 Helmo,M., Krohn,K.J., Laliochi,M.D., Mullis,P.E., Antonarakis,S.E.,
 Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N.
 Mutations in a novel zinc finger protein AIR are responsible for
 autoimmune polyglandular disease type I (APECED)
 JOURNAL Unpublished
 TITLE Location/Qualifiers
 FEATURES
 source 1..200
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="q22.3"
 exon 1..200
 /note="trapped exon"
 BASE COUNT 37 a 48 c 68 g 35 t 12 others
 ORIGIN
 Query Match 7.6%; Score 170.8; DB 17; Length 200;
 Best Local Similarity 90.7%; Pred. No. 1e-18;
 Matches 175; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2177 AAAAACCACATACATTAATTATCATCTCGACACTT 2214
 |||| | | | | | | | | | | | | | | | | | |
 Db 44 AAAAAAAAAAAAAAAAACTTATGATACCTCGACCT 7

RESULT 7
 A0425335 620 bp DNA linear GSS 24-MAR-1999
 LOCUS A0425335
 DEFINITION DNA sequence.
 CITEB1-E1 Homo sapiens genomic clone 2572112,
 A0425335
 VERSION A0425335.1 GI:4498557
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 620)
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
 Venter,J.C.
 TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1997)
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..620
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2572112"
 /clone_lib="CITBI-E1"
 /sex="male"
 /cell_type="sperm"
 /note="Vector: pBelOBAC11, Site_1: EcoRI; Site_2: EcoRI;
 Caltech Human BAC Library D"

BASE COUNT 149 a 178 c 117 g 176 t
 ORIGIN

Query Match 6.9%; Score 153.8; DB 17; Length 620;
 Best Local Similarity 82.6%; Pred. No. 4.6e-16;
 Matches 176; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1996 CTAATAAATATATAATAGCTGGTGTGTGGTGGTGTATCCAGCTACAT 2055
 ||||| | | | | | | | | | | | | | | | | | |
 Db 516 CTACTATAAATACAAAAATAGCTGGTGTGTGGTGGTGTATCCAGCTACTT 457

QY 2056 GGGAGCTTGAAGCATAGAACTCACTTGAAGTGGAGGTTGAGTCACTGAGT 2115
 ||||| | | | | | | | | | | | | | | | | | |
 Db 456 GGGAGCTTGAAGCATAGAACTCACTTGAAGTGGAGGTTGAGTCACTGAGT 397

QY 2116 TTGGCGCACTGCAGCTCAGTCTGGCGCAAGAGTGAATCCGCTCAAAAAACAAACA 2175
 |||| | | | | | | | | | | | | | | | | | |
 Db 396 TTGCACACATCGCATCTCACTGCTGGCGCAAGAGTGAATCCATCCAAAAA 337

QY 2176 AAAAAACACATACATTAATTTATCATCTCGA 2208
 || | | | | | | | | | | | | | | | | |
 Db 336 AATTATGAAATTGGCTATATTAGGTTCTCCA 304

RESULT 8
 A0338901 516 bp DNA linear GSS 12-JAN-1999
 LOCUS A0338901
 DEFINITION HS_2215_B2_B08_MR CIT Approved Human Genomic Sperm Library D Homo

ACCESSION A0338901
 VERSION A0338901.1 GI:4144917
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 516)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2215 row: D column: 16
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 516.

FEATURES
 source
 Location/Qualifiers
 1..516
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-2215 Col=16 Row=D"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
 E-Coli DH10B"
 /sex="male"

BASE COUNT 122 a 135 c 96 g 163 t
 ORIGIN

Query Match 6.8%; Score 152.8; DB 17; Length 516;
 Best Local Similarity 82.5%; Pred. No. 7.1e-16;
 Matches 175; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1994 ACCTCTAATAAATATAATAGCTGGTGTGTGGTGGTGTATCCAGCTAC 2053
 | | | | | | | | | | | | | | | | | | | | | |
 Db 329 ATCTACTATAAATATCAAAAAATTTGGTGTGTGGTGGTGTATTCAGCTAC 270

QY 2054 ATGGAGCCTGAGCATGGAATCACTTGAACCTGGAGGTGGAGTTCAGTGAAGCTGA 2113
 ||||| | | | | | | | | | | | | | | | | | |
 Db 269 TTGGGAGCTGAGCGCAGGAAGAACGCTTGAACCCAGAGGCAAGAGTTCAGTGAAGCCGA 210

QY 2114 GATTGGCCACTGCAGCTCAGTCTGGTGGCGCAAGAGTGAAGTCCGCTCAAAAAACAAA 2173
 || | | | | | | | | | | | | | | | | | | | | |
 Db 209 GATCGCGCCTACCTGCACTCCAGCTGGCGCAAGAGTGAATTCGCTCAAAAAA 150

QY 2174 CAAAAAACCATATACATTAATTTATCATCT 2205
 || | | | | | | | | | | | | | | | | |
 Db 149 AAAAAAATGCACATATATTAATTTATTTT 118

RESULT 9
 A1889995 546 bp mRNA linear EST 07-MAR-2000
 LOCUS A1889995
 DEFINITION wmb0c03.x1 NC1_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2442244 3',
 similar to contains Alu repetitive element; contains element THR
 repetitive element ;, mRNA sequence.
 ACCESSION A1889995
 VERSION A1889995.1 GI:5595159
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.

contact: shayling.zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research

... can be sequenced by

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2003, 11:38:50 ; Search time 5652.5 Seconds

(without alignments)
11558.750 Million cell updates/sec

Title: US-09-509-595B-1

Perfect score: 2245

Sequence: 1 cggcgccagccgcccgcg.....ggcattccatcctcatgtaa 2245

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenBml: *
1: gb_da: *
2: gb_hcg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rod: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_sy: *
39: em_hlgo_hum: *
40: em_hlgo_mus: *
41: em_hlgo_other: *

Pred. No. is the number of results predicted by chance to have a

Score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2245	100.0	2245	6 A97284	A97284 Sequence 1
2	2245	100.0	2245	9 HSAPCED	Z97990 Homo Sapien
3	2018.4	89.9	2027	9 AB006682	AB006682 Homo sapi
4	1025.2	45.7	1545	9 AB006683	AB006683 Homo sapi
5	964.2	42.9	1636	10 AF128772	AF128772 Mus muscu
6	964.2	42.9	1636	10 AF128773	AF128773 Mus muscu
7	964.2	42.9	1636	6 A97293	A97293 Sequence 10
8	964.2	42.9	1636	10 MMU243821	AJ132243 Mus muscu
9	964.2	42.9	1636	10 AF079536	AJ243821 Mus muscu
10	964.2	42.9	1636	10 AF079536	AF079536 Mus muscu
11	948.2	42.2	1933	10 AF128115	AF128115 Mus muscu
12	939.8	41.9	1924	10 AF128116	AF128116 Mus muscu
13	923.8	41.1	1921	10 AF128117	AF128117 Mus muscu
14	851.2	37.9	1463	9 AB006685	AB006685 Homo sapi
15	737.6	32.9	1637	10 AF128122	AF128122 Mus muscu
16	736.6	32.8	1759	10 AF128118	AF128118 Mus muscu
17	721.6	32.1	1634	10 AF128123	AF128123 Mus muscu
18	720.6	31.8	1756	10 AF128119	AF128119 Mus muscu
19	713.2	31.7	1625	10 AF128124	AF128124 Mus muscu
20	712.2	31.7	1747	10 AF128120	AF128120 Mus muscu
21	697.2	31.1	1622	10 AF128125	AF128125 Mus muscu
22	696.2	31.0	1744	10 AF128121	AF128121 Mus muscu
23	546	24.3	2586	9 HSTRPER	X80853 H. sapiens g
24	546	24.3	20000	9 AB006684	AB006684 Homo sapi
25	546	24.3	42133	9 AP001060	AP001060 Homo sapi
26	546	24.3	34000	9 AP001754	AP001754 Homo sapi
27	544.4	24.2	36284	9 HSA9610	AJ009610 Homo sapi
28	526	23.4	110000	2 AC003656_5	Continuation (6 of
29	170.8	7.6	200	17 HSDML04M6	AJ001114 Homo sapi
30	167.8	7.5	9876	6 AX344995	AX344995 Sequence
31	159.2	7.1	161361	9 AC090939	AC090939 Homo sapi
32	159.2	7.1	172676	9 AC015711	AC015711 Homo sapi
33	158.2	7.0	178152	9 AC096633	AC096633 Homo sapi
34	158.2	7.0	209216	9 AF190464	AF190464 Homo sapi
35	158	7.0	192007	2 AC090061	AC090061 Homo sapi
36	157.6	7.0	191872	2 AC129837	AC129837 Papio cyn
37	157.6	7.0	202414	2 AC090966	AC090966 Papio cyn
38	157.2	7.0	138292	9 AC015550	AC015550 Homo sapi
39	157.2	7.0	204654	9 CNG0507C9	AL355075 Human chr
40	157	7.0	85446	9 AC090013	AC090013 Homo sapi
41	157	7.0	169103	2 AC025687	AC025687 Homo sapi
42	156.8	7.0	111234	9 AC013451	AC013451 Homo sapi
43	156	6.9	72651	9 AP001855	AP001855 Homo sapi
44	156	6.9	135424	2 AC069040	AC069040 Homo sapi
45	156	6.9	139931	9 AC078818	AC078818 Homo sapi

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
A97284	A97284	Sequence 1 from Patent WO918197.	A97284	A97284.1	GI:6780668	unidentified	unclassified.	1 (bases 1 to 2245)	Yaspo,M. and Lehrach,H.	NUCLEIC ACID MOLECULE ENCODING A (POLY)PEPTIDE CO-SEGREGATING IN MUTATED FORM WITH AUTOIMMUNE POLYENDOCRINOPATHY CANDIDIASIS ECTODERMAL DYSTROPHY (APECD)
A97284	A97284	Sequence 1 from Patent WO918197.	A97284	A97284.1	GI:6780668	unidentified	unclassified.	1 (bases 1 to 2245)	Yaspo,M. and Lehrach,H.	NUCLEIC ACID MOLECULE ENCODING A (POLY)PEPTIDE CO-SEGREGATING IN MUTATED FORM WITH AUTOIMMUNE POLYENDOCRINOPATHY CANDIDIASIS ECTODERMAL DYSTROPHY (APECD)

JOURNAL Patent: WO 918197-A 1 15-APR-1999;
 MAX PLANCK GESELLSCHAFT (DE); YASPO MARIE LAURE (DE)
 FEATURES Location/Qualifiers
 source 1..2245
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 419 a 765 c 697 g 364 t
 ORIGIN

Query Match 100.0%; Score 2245; DB 6; Length 2245;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGCGACAGCGGGGGGAGGGCCGACAGCCCGCGGGAGCCCGGAGGCGCAAGAGAGG 60
 Db 1 CGGGCGACAGCGGGGGGAGGGCCGACAGCCCGCGGGAGCCCGGAGGCGCAAGAGAGG 60
 QY 61 GCTGCGAGTGTCCGGGAGCCAGCCGCTCCGCCAGCCCGGGGTCGCCCGCCACGCC 120
 Db 61 GCTGCGAGTGTCCGGGAGCCAGCCGCTCCGCCAGCCCGGGGTCGCCCGCCACGCC 120
 QY 121 ATGGCGACGAGCGGGGCGCTACGCCGCTTGTAGGCTGCACCGCAGAGATCGGGGTG 180
 Db 121 ATGGCGACGAGCGGGGCGCTACGCCGCTTGTAGGCTGCACCGCAGAGATCGGGGTG 180
 QY 181 GCGGTGACAGCGGCTTCCCACTGCTGACGCGCTGCTGACAGAGAGTGTCCCGAG 240
 Db 181 GCGGTGACAGCGGCTTCCCACTGCTGACGCGCTGCTGACAGAGAGTGTCCCGAG 240
 QY 241 GACAGAGTTTACGAGAGAGCGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 Db 241 GACAGAGTTTACGAGAGAGCGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 QY 301 GCGCTCTGTCTGTGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 Db 301 GCGCTCTGTCTGTGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 QY 361 CTGTTCAAGAGACTCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 Db 361 CTGTTCAAGAGACTCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 QY 421 CCCAAGATGTGACCTCAGCCAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 Db 421 CCCAAGATGTGACCTCAGCCAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 481 GCTTTGTACCGGACCCAGACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 Db 481 GCTTTGTACCGGACCCAGACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 541 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 Db 541 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 601 AAGCCCCCAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 Db 601 AAGCCCCCAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 661 ATTCAACCATGTACCTTCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 Db 661 ATTCAACCATGTACCTTCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 721 GAGAGCGGAGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 Db 721 GAGAGCGGAGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 QY 781 AAGAGTGCATCCAGGTGTGGTGGAGATTCTACACTCCAGAGAGAGAGAGAGAGAG 840
 Db 781 AAGAGTGCATCCAGGTGTGGTGGAGATTCTACACTCCAGAGAGAGAGAGAGAGAG 840
 QY 841 AGTGGAG 900
 Db 841 AGTGGAG 900

QY 901 GCCAGAGGCGCTGCCCGGTGAGAGTGAAGCTGAGGCGCAGAGAGAGAGAGAGAG 960
 Db 901 GCCAGAGGCGCTGCCCGGTGAGAGTGAAGCTGAGGCGCAGAGAGAGAGAGAGAG 960
 QY 961 GCCCTGTGGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 Db 961 GCCCTGTGGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1021 GTGTGTGAG 1080
 Db 1021 GTGTGTGAG 1080
 QY 1081 GCGT 1140
 Db 1081 GCGT 1140
 QY 1141 CTGAG 1200
 Db 1141 CTGAG 1200
 QY 1201 CCGGTGAG 1260
 Db 1201 CCGGTGAG 1260
 QY 1261 CCGGTGAG 1320
 Db 1261 CCGGTGAG 1320
 QY 1321 CCGGTGAG 1380
 Db 1321 CCGGTGAG 1380
 QY 1381 GGTCTGAG 1440
 Db 1381 GGTCTGAG 1440
 QY 1441 AGGAG 1500
 Db 1441 AGGAG 1500
 QY 1501 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 Db 1501 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1561 CAGAGCCCTGTGAG 1620
 Db 1561 CAGAGCCCTGTGAG 1620
 QY 1621 AAGAGATGACAG 1680
 Db 1621 AAGAGATGACAG 1680
 QY 1681 AAGAGATGACAG 1740
 Db 1681 AAGAGATGACAG 1740
 QY 1741 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 Db 1741 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 QY 1801 AG 1860
 Db 1801 AG 1860
 QY 1861 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
 Db 1861 GCGGCGACAGCGGCTGACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
 QY 1921 CATGTGCTGAG 1980
 Db 1921 CATGTGCTGAG 1980
 QY 1981 TCCTTGCTGAG 2040

Qy	901	GGCCAGGAGCGCTGGCCCCGGGGGAGAGTGAAGCTTAGCTGAGCCAGACAGAGCTTCCC	960
Db	901	GGCCAGGAGCGCTGGCCCCGGGGGAGAGTGAAGCTTAGCTGAGCCAGACAGAGCTTCCC	960
Qy	961	GGCCCTTGTGAGCCCTCCCAAGTGAACCCCAAGTCTCCACAAGAAATGAAGAGAGTGTCC	1020
Db	961	GGCCCTTGTGAGCCCTCCCAAGTGAACCCCAAGTCTCCACAAGAAATGAAGAGAGTGTCC	1020
Qy	1021	GTTGTGTGGGAGCGGGGGAGCTCATCTGCTGTAGCGCTGTCCCTCGGGCTTCCACTGT	1080
Db	1021	GTTGTGTGGGAGCGGGGGAGCTCATCTGCTGTAGCGCTGTCCCTCGGGCTTCCACTGT	1080
Qy	1081	GGCTGGCTGTCCCTCCGGCTCCGGGAATCCCAAGTGGACCTGGAGTCTCCAGCTGC	1140
Db	1081	GGCTGGCTGTCCCTCCGGCTCCGGGAATCCCAAGTGGACCTGGAGTCTCCAGCTGC	1140
Qy	1141	CTGCAGCAACAGTTCAGAGAGTGCACCCCGGCGACAGAGGCGCCCGGCCCCAGAGCCA	1200
Db	1141	CTGCAGCAACAGTTCAGAGAGTGCACCCCGGCGACAGAGGCGCCCGGCCCCAGAGCCA	1200
Qy	1201	CCCGTGGAGACCCCGCTCCCCCGGGCTTAAAGTGGCGGGAGAGAGATTAAGAGTCCA	1260
Db	1201	CCCGTGGAGACCCCGCTCCCCCGGGCTTAAAGTGGCGGGAGAGAGATTAAGAGTCCA	1260
Qy	1261	CTGTGGGGAACCCCTAGCGGGCATGTGACACACTTCTTCAACAGACACTCCGCGCTCG	1320
Db	1261	CTGTGGGGAACCCCTAGCGGGCATGTGACACACTTCTTCAACAGACACTCCGCGCTCG	1320
Qy	1321	CTTCTGCACGCCCCGCTGCCAGGGGCTGGACTCCCGGGCCCTGCACCCCTCACTGTGTGTG	1380
Db	1321	CTTCTGCACGCCCCGCTGCCAGGGGCTGGACTCCCGGGCCCTGCACCCCTCACTGTGTGTG	1380
Qy	1381	GGTCTGAGGGGTAGCAGAAACCTGGCTCTGTGTGCGGTTGCGGGGAGTTCGAGATGTGT	1440
Db	1381	GGTCTGAGGGGTAGCAGAAACCTGGCTCTGTGTGCGGTTGCGGGGAGTTCGAGATGTGT	1440
Qy	1441	ACGAGAGTGTGGGGGTGTACTACTGTGCGCGCTGTCCACTGTGCGGCTCCACTTCCA	1500
Db	1441	ACGAGAGTGTGGGGGTGTACTACTGTGCGCGCTGTCCACTGTGCGGCTCCACTTCCA	1500
Qy	1501	GGCGGACACTCCCGGCGCGGGAGCGGGGCTGGAGATCCGTCTCAGAGACAGTGAAC	1560
Db	1501	GGCGGACACTCCCGGCGCGGGAGCGGGGCTGGAGATCCGTCTCAGAGACAGTGAAC	1560
Qy	1561	CGAGCCCCGTGTGAGAGGGGTGTGTGACCCAGCCAGCCCGCGCTGAGCCCTTGAGCCGTCC	1620
Db	1561	CGAGCCCCGTGTGAGAGGGGTGTGTGACCCAGCCAGCCCGCGCTGAGCCCTTGAGCCGTCC	1620
Qy	1621	AAGGATGACACTGTCCAGTACAGACGCCGCTGTGCACAGGAGTACCTGTGACCTTCTG	1680
Db	1621	AAGGATGACACTGTCCAGTACAGACGCCGCTGTGCACAGGAGTACCTGTGACCTTCTG	1680
Qy	1681	AAGGAGACACCTTCATGATGGATCTCGACGTGGGGCATCCAGAGCATGGCCGCTCCGGG	1740
Db	1681	AAGGAGACACCTTCATGATGGATCTCGACGTGGGGCATCCAGAGCATGGCCGCTCCGGG	1740
Qy	1741	GGCCCCCTTCCCTCTGACCCCAAGATGAGCGGGACATGTACGCTCTATGAGAGAGTGTG	1800
Db	1741	GGCCCCCTTCCCTCTGACCCCAAGATGAGCGGGACATGTACGCTCTATGAGAGAGTGTG	1800
Qy	1801	AGAGAGACACTTCCTTCTCAGTCTGTGAAGCGCGGCTGGGATCAAGAAAGGGAGACAG	1860
Db	1801	AGAGAGACACTTCCTTCTCAGTCTGTGAAGCGCGGCTGGGATCAAGAAAGGGAGACAG	1860
Qy	1861	GGCCACCTTGTTCAGTGTCTGGCTGTAAACAGCTCTGTGTCTGTGGGACACACAGCAT	1920
Db	1861	GGCCACCTTGTTCAGTGTCTGGCTGTAAACAGCTCTGTGTCTGTGGGACACACAGCAT	1920
Qy	1921	CATGTGCTGTGAATTAACCCCTGGCCCACTTCTGTACTGTGAGATCCCGCGGAGCCCTC	1980
Db	1921	CATGTGCTGTGAATTAACCCCTGGCCCACTTCTGTACTGTGAGATCCCGCGGAGCCCTC	1980

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 1656) Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q., Dettler,J.C. and She,J.X.
TITLE	Chromosomal localization and complete genomic sequence of the murine autoimmune regulator gene (Aire)
JOURNAL	Autoimmunity 31 (1), 47-53 (1998)
MEDLINE	20059112
PUBMED	10593569
REFERENCE	2 (bases 1 to 1656) Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q., Dettler,J.C., Davoodi-Semimi,A. and She,J.X.
AUTHORS	Complete genomic sequence, gene structure and localization of the mouse Aire gene
TITLE	unpublished
JOURNAL	3 (bases 1 to 1656) Ruan,Q.G., Wang C.Y., Shi,J.D. and She,J.X.
REFERENCE	Direct Submission
AUTHORS	Submitted (17-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road

FEATURES	Location/Qualifiers
source	1. .1656

```

/organism="Mus musculus"
/strain="SJL"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1656
/gene="Aire"
1..1656
DS

```

BASE COUNT	344 a	532 c	479 g	301 t
ORIGIN				

Query Match	42.9%;	Score 964.2;	DB 10;	Length 1656;
Best Local Similarity	76.0%;	Pred. No. 1.7e-160;		
Matches 1249;	Conservative	0;	Mismatches 373;	Indels 21; Gaps 4;

OY	129	GGACCGGGGGCTACGCCGGGCTTCGTGAGGCTGCACCGCAGGAGATCGCGTGGCCGTGGA	188
Db	12	GGATGGAAATGCTACGCCCGCTCTCTCTAGGCTGCACCGCACCGAGATCGGGGTGGCCATAG	71
OY	189	CAGCCCTTTCCACTGCTGTGCACGCGCTGGCTGACACGACGTGTCCTCCGAGGACAACTT	248
Db	72	CAGTCCCTTTCCGCTGCTGCATGCTCTGACCGCACCCACAGCGTGCCTCTGAGGACAACTT	131
OY	249	TCAGAGACGCGTTCCTCTGGAAGAAAAGAGGGGCGCCCGACGGCTTCACAGCCCTCTCT	308
Db	132	CCAGSAGACGCTCCCTCTCTGAAAGGAGAAAGAAAGGCTGGCCCGCCAGAGGCTTTCACAGCCCTCTCT	191
OY	309	GTCCTGCTGCTGACCCAGGACTCCACAGCCATCTGTGACTTCGTGAGAGTGCTTTCAA	368
Db	192	GTCCTGCGCTCTGACCCCGGACAGTGGGGCCCATCTCGATTCTGTGAGAGATTCTTTAA	251
OY	369	GGACTACAACTGGAGGCGCTATTGCGGGCTGTACAGCCCATCTCTGGACAGCTTCCCAAGA	428
Db	252	GGACTACAAATGTGGAGCGCGTACAGCCCGCTCTCATAGCATCTCTGGACAGGCTTCCCAAGA	311

QY	429	TTGTGGACCTCAGACAGCCCGGAAAGGGAGGAGACCCCCCGCGCTGCCCAAGGCTTTTGGT	488
Db	312	TTGTGGACCTTAAACCAAGTCCCGGAAAGGAGAAACCCCTTGTGCTCCAAAGCCGCGGT	371
QY	489	ACCGCACCCACACACTCCCCACCAAGAGAGAAAGCCCTCAAGAAAGAGCTTCAGACTCCGCGCC	548
Db	372	ACTGCCACCCAGAACCCCCACCACAGAGAAAGACACTGGAGAGAGCTTGACACACCCACC	431
QY	549	AGCAGCCCTGACTCCAGGGGACACCGCCAGGCTCTCAACTGAAGGCCAAGCCCC	608
Db	432	AGCAACTCTGGGCTTCAAAAGACGCTCCAGCCCGAGGCTCCACCTGAAGACTTAAGCCCC	491
QY	609	CAAGAAAGCCGAGACACACGCGAGAGACAGACAGCCCTTCACCTGGGAGACGGATTACAC	668
Db	492	TAAAGAGCCAGATGGCCAACTTGGAGTACACACACACTCTCTTGGAAACGGAATTTACAG	551
QY	729	AGGGCCCTGGAGGGGATCTCATCCAGCAGGTTTGAAGTACAGCGCGCTCCAAAGATG	788
Db	612	AGGGCCCTGGAGAGGATCTTATCCAGACAGTGTTTGAGTCAAGAAATCCAAAGATG	671
QY	789	CATCCAGGTTGGTGGGAGTTCTACACTCCAGCAAGTTGAAGACTCC---GGCAGTGG	845
Db	672	CATTCAAGTTGGGGAGAGTTTATACACCCACCAACTTTGAAAGACCCCGAGTCAATT	731
QY	846	GAAAGAACAGAGCCCGCAGACACACATGGCCCGCAACCCCTGTTCGAGCCAAAGGACCCCA	905
Db	732	GAAAGAACAGAGCCCGCGAGATGGTACACACTTAAGCCAGTGTTCGAGAACAGGAGACCCA	791
QY	906	GGGCGCTGCCCCGGTGGAGGTGAAGCTTAGGCTGGGCCACAGGCGACTTCCCGCCC	965
Db	792	GGTCACTATTACTGGTAGAGATGAGCAGAAAGTGGGCCACAGTGTGGGGTTCTCCCT	851
QY	966	TCTGGCCCTTCCCCAGATGACCCCGACGCTCCACAGAGATGAGACGAGTGTCCGTGTG	1022
Db	852	TCCATCCCTCCCGAGTGGAGCCCGAGTTTAAACAGAAAGACGAGATGAGTGTCCGTGTG	911
QY	1026	TGCGGAGCGCGGGAGGCTATCTCTGTGAAGGCTGGCCCGGGCCCTCCACTGGCCCTG	1082
Db	912	CCAGCAGCGGAGGTAGCTCATCTTGTGTGACGGTGTGCCCGGGCCCTCCACTGGCCCTG	971
QY	1086	CCTGTCCCTCCGCTCCGCGGAGATCCCCAGTGGGACTGGAGGTGTCTCCAGCTGCCTGCA	1144
Db	972	CCTGTCCCACTCTGTGCAGAGAGATCCCAATGGCGCTGTGAGATGTCTCTGCTGCTCCA	1033
QY	1146	GGCAACAGTCCAGAGAGTGCACGCCCGGGCAGAGAGACCCCGGCCACAGACCCAGT	1205
Db	1032	GGGAGAGTCCACAGAAAGCTGTCCAGCTGAGGTGTCCAGGCTCCCGGAGCTACTGC	1092
QY	1206	GGAGACCCCGCTCCCGCGGGGGTGTAGTGGGGGAGAGAGAGTAAAGGTTCACACTGG	1266
Db	1092	AGAGACCCCATCTCTGTGGAGCTGAGAGTGAAGTTCAAGAAAAACCAAGGGGCCACTCAG	1151
QY	1266	GGAACCCCTAAGCCCGCATGTGAGACAGACTCTTGTCTAAAGACACTCCCGCTCCGCTTC	1322
Db	1152	GGAGCTCAAAAGCCAGCTCTGATGCTGTGTGCACATATGTGAACCTGTCTGGCCCGCACCC	1211
QY	1326	TGCAGCCCCCTGCGCAGGAGCTTGACATCTCTGGCCCTGCACACCCCTCACTGTGTGGGTGC	1385
Db	1212	TGCAGCTCTCTCT-----GCTGGAGCCCTTCAGCAGTGTCCCTCTCAATGAGATGGGAA	1265
QY	1386	TGAGAGGTACAGAAAGCTGTGCTCTGGTGTGGCTGTGCGGGGTGTGCGAGATGTGTACGGA	1445
Db	1266	TGAGAGGCGGCACAGTTCACACACAAAGCGGCGCATCAATGTGTGTGGCGATGTGACGCA	1329
QY	1446	CTGTGCTGGGTACTACATGACGCGCGCTGCTTCCACTGGGCTGCGACTTCCAGCGCGG	1505
Db	1326	GGTGTGGGTGTGACACAGTGTGCCCTGCTTCACACTGTGGCGCTGCACCTTCCAGCGCC	1385
QY	1506	CACCTCCCGGCGGGAGCGGGCTGTGCGTGCAGATCTTGCTGAGAGAGTG-----AC	1559

Db 1386 CCCCCCGGGGGGACCAATCTCCGTCGAAATCTGCTGACAGACTCCAGCTCCAC 1445
 QY 1560 CCGAGCCCTTGAGAGGGGCTGCTGCGCCCGCCCGCCCGCTGCGCCCTGCGCTGC 1619
 Db 1446 GCCAGGACACCGGGCGAAGCTGTACCACTCTGCGCCCGCTGCGAGCTGGCTTC 1505
 QY 1620 CAA-----GATGACACTGCGACGTACAGAGCCCGCTGTCACAGAGATGATGAGTC 1673
 Db 1506 CAAAGTAGGGGACGACTGTGCTAGTACGACCGCTGTTACATAGGAGAGCTGAGTC 1565
 QY 1674 CCTCTGAGGACGACACTCTGATGATGATCTGATGATGATGATGATGATGATGATGATG 1733
 Db 1566 CTTCTCATAGTACGACTATTGAGAGGATCTGATGATGATGATGATGATGATGATGATG 1625
 QY 1734 TCCGCGGCGCCCTTCCCTCTCT 1756
 Db 1626 CCGCTGCGCGAGACACACCT 1648
 RESULT 6
 AF128773 1656 bp mRNA linear ROD 26-OCT-2001
 LOCUS AF128773 Mus musculus strain NOD autoimmune regulator (Aire) mRNA, partial
 DEFINITION cds
 ACCESSION AF128773 GI:7108574
 VERSION AF128773.1
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1656)
 AUTHORS Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q., Dettler,J.C. and She,J.X.
 TITLE Chromosomal localization and complete genomic sequence of the murine autoimmune regulator gene (Aire)
 JOURNAL Autoimmunity 31 (1), 47-53 (1999)
 MEDLINE 20059142
 PUBMED 10593569
 REFERENCE 2 (bases 1 to 1656)
 AUTHORS Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q., Dettler,J.C., Davoodi-Sermlomi,A. and She,J.X.
 TITLE Complete genomic sequence, gene structure and localization of the mouse Aire gene
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1656)
 AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
 TITLE Direct Submission
 JOURNAL Submitted (17-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road
 FEATURES
 source Location/Qualifiers
 1..1656
 /organism="Mus musculus"
 /strain="NOD"
 /db_xref="taxon:10090"
 /chromosome="10"
 /map="between D10Mit31 and D10Mit10"
 1..>1656
 /gene="Aire"
 1..>1656
 /gene="Aire"
 /function="transcription factor"
 /codon_start=1
 /product="autoimmune regulator"
 /protein_id="AAF36482.1"
 /db_xref="GI:7108574"
 /translation="MAGDGMRLRLRLHRTIEAIVASAPPLLAHLADHDVVPEDKE
 QETLRKEKESGPOAFHALSLWLTROSGALIDFWILFKOVNLEYSRLSHLSDGFP
 KDVLNOSRKRCPLAGKAVLPRPTKALPEPATLASSVSPGSHLK
 TRPKKPDGNESOHLEPLNGIQTMAASVQAVAVASDVGRTGAVSGILIQVEFS
 GRSKCKIQVGSEFTTPNKFEDPSGLNKKRASGSLKPVNRAKGAQVTPGRDEKVG
 QCGVPLPLSPSEPOYNQKNEDCAVCHDGEILICDGCPRAFHLACISPLDGEIPS

GIMRSCCLOGRYOQNLSPREVSRLPAETPLVLRASERTRBPSREKASSDA
 AVTYVILAFHPAPLIEBSALCPILSAGNEGRPAIPASRCSYCGTTEVLRCAHCA
 AAFHMRCHPEPTAARPCDTNLRKSCSDSTPTPTDTPDEAVPTSGPRAPIAKVGDOS
 ASHDPVLRHDELSELNHSFDDILQWAIQISMSRPLAETPEFS"
 BASE COUNT 344 a 532 c 479 g 301 t
 ORIGIN
 Query Match 42.9%; Score 964.2; DB 10; Length 1656;
 Best Local Similarity 76.0%; Pred. No. 1,7e-160;
 Matches 1249; Conservative 0; Mismatches 373; Indels 21; Gaps 4;
 QY 129 GAGAGCGGCGCTACCGCGGCTTTCAGGCTGACGACGAGGAGTCCGGGCGCTGGA 188
 Db 12 GATATGGATCTACCGCGCTGCTGAGAGCTGACCGACGACGATCGCGGCTATGA 71
 QY 189 CAGCGCTTCCCACTGCTGACGCGCTGCTGACGACGAGCTGCTCCCGAGGACAAATT 248
 Db 72 CAGTGCCTTCCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
 QY 249 TCAGAGAGCGCTTCATCTGAAGAAAGAGGCTGCCCGCCCTTCCAGCGCTCTCT 308
 Db 132 CAGGAGAGCGCTCTGTCGAAGAGAGAGAGGCTGCCCGCCAGCGCTTCCAGCGCTCT 191
 QY 309 GTTCCTGCTGCTGACCGACGACGACGACGACGACGACGACGACGACGACGACGACG 368
 Db 192 GTCTGCTGCTGCTGACCGCGGACGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCT 251
 QY 369 GGACTACAACTGAGAGCGCTATGCGCGCTGACGACGACGACGACGACGACGACGAC 428
 Db 252 GGACTACAACTGAGAGCGCTGACGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 311
 QY 429 TGTGAGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 488
 Db 312 TGTGAGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 371
 QY 489 ACCGCGACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 548
 Db 372 ACTGCGACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 431
 QY 549 AGCAGCGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 608
 Db 432 AGCAGCGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 491
 QY 609 CAAAGCGCGGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 668
 Db 492 TAAAGCGCGAGTGGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 551
 QY 669 CATGTCAGCTTCAGTCCAGAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 728
 Db 552 CATGTCAGCTTCGTCAGAGAGCTGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 611
 QY 729 AGGGCGCTGAGAGGAGATCTCATCCAGAGAGTGTGAGTGTGAGGCGCTCCAGAGAGT 788
 Db 612 AGGGCGCTGAGAGGAGATCTCATCCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 671
 QY 789 CATCAGCTTGTGTGGAGTCTTACCTCCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAG 845
 Db 672 CATTCAGCTTGTGTGGAGTCTTATACACCAAGTGTGAGTGTGAGTGTGAGTGTGAG 731
 QY 846 GAAAGACAGGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 905
 Db 732 GAAAGACAGGCGCGGAGTGTGACGCTTAAAGCAAGTGTGAGTGTGAGTGTGAGTGTG 791
 QY 906 GGGCGCTGCGCGCGCTGAGAGTGTGAGGCTGAGGCGGAGGAGGAGGAGGAGGAGGAG 965
 Db 792 GGTACATATCTGTGTAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 851
 QY 966 TGTGCGCTTCCCGAGTACCGGACGCTCCAGCAAGATGAGAGAGTGTGAGTGTGAGTGT 1025
 Db 852 TCAATCCCTCCCGAGTACCGGAGCGGTTAACCAAGAGAGTGTGAGTGTGAGTGTGAG 911
 QY 1026 TCGGAGCGGCGGAGCTCATCTGTGACGCTGCTGCGGCGCTTCCAGCTGCGCTG 1085

QY	1146	GGCAACAGTCACAGAGGTGACGCCCGGACAGAGAGACCCCGCCAGACACCCTT	1205
Db	1032	GGGACAGATTCACACAGAACTGTGCCAGCTTGAGGTGTCCAGGACCCCGGAGCTTACCTGC	1091
QY	1206	GGAGACCCCGCTCCCGCGGGGCTTAGTGGGGGAGAGAGAGATTAAGAGTGCACCTGG	1265
Db	1092	AGAGACCCCGCATCTGTGTGGACTGAGGTAGCTTCAGAGAAAACAGGGGCCCATTCAG	1151
QY	1266	GGAAACCCCTTAGCCGGCATGGACACAGACTTGTCTACAAAGACCTCCCGGCTCCGGCTTC	1325
Db	1152	GGAGCTCAAAAGCCAGCTCTGATGCTCTGTGCATATATGTAACCTGCTGGACCCCGACCC	1211
QY	1326	TGAGACCCCGCTCCAGAGGCTGGAGTCTCTGGCCGACACCCCTACTGTGTGGGTCC	1385
Db	1212	TGCAAGCTCTCT-----GCTTGAAGCTTACAGCATGTGTGCTCTTACTGTAGTCTTGGAA	1265
QY	1386	TGAGGGTCAAGACAGACCTGGCTCTGTTGGGCGCTTGGCGAGATGTATGTAACGGA	1445
Db	1266	TGAGGGGCGGCGCAGGTTCAGACACCAAGCGGGCATTCAGTGTGTGTGGCATGGACCGA	1325
QY	1446	CTGTCTGGCGTGTATCTACTGCGCGCTGCTTCCATGAGCGCTGCCACTTCCAGCCGG	1505
Db	1326	GGGTGTGGCGTGTGACACACTGTGCCCTGCTTCCACTGAGCGCTGCCACTTCCGACGGC	1385
QY	1506	CACCTCCGCGCGGACGAGGAGCTGGCGCTGAGATCTCTGACAGAGAGTGTG-----AC	1559
Db	1386	CGCGCGCGCGCGCGGACCAATCTCCGCTGTCAATCTCTGTGCAAGACTTGCATCCAC	1445
QY	1560	CCGAGCCCTGTGTGAGGGGTGCTGGCCCCCAGCCCGCGCGCTTGCGCTGGGCTGAC	1619
Db	1446	GGCAGGCGACACGGGGGAGAGCTGTACCCACTCTGTGGGCGCCGCGACACTGGGCTTGC	1505
QY	1620	CAA-----GGATGACACAGCGCAGTACAGAGCCCGCTGTGACAGGATGACCTGGAGTC	1673
Db	1506	CAGAGTAGGGGAGACATCTGTCTAGTACAGACCTGTCTTACATAGGAGAGACCTGGAGTC	1565
QY	1674	CCTTCTGAGCGAGACACCTTGGATGGCATCTGTCAGTGGGCGATCCAGACAGATGGCCCG	1733
Db	1566	CCCTCTCATATGAGCACTCATTTGAAGGCACTCTGCAAGTGGGCGATCCAGACATGTACAG	1625
QY	1734	TCCGGCGGCGCCCTTCCCTCCT 1756	
Db	1626	CCCGCTGGCGGAGACACACCT 1648	
RESULT 8	MMU132243		
LOCUS	MMU132243	1659 bp	linear
DEFINITION	Mus musculus mRNA for Aire protein.		ROD 24-APR-1999
ACCESSION	AJ132243		
VERSION	AJ132243.1	GI:4456674	
KEYWORDS	Aire gene; Aire protein.		
SOURCE	Mus musculus.		
ORGANISM	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1659)		
AUTHORS	Bleachschmidt, K., Schweiger, M., Wertz, K., Poulson, R., Christensen, H.M., Rosenthal, A., Lehrach, H. and Yaspo, M. L.		
TITLE	The mouse Aire gene: comparative genomic sequencing, gene organization, and expression		
JOURNAL	Genome Res. 9 (2), 158-166 (1999)		
MEDLINE	99148139		
PUBMED	10022980		
REFERENCE	2 (bases 1 to 1659)		
AUTHORS	Karin, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-JAN-1999) Karin B., Max Planck Institute for Molecular Genetics, Ihnestrasse 73, Berlin D-14195, Germany		
FEATURES	Location/Qualifiers		
SOURCE	1. 1659		
	/organism="Mus musculus"		

```

/strain="129"
/db.xref="taxon:10090"
/chromosome="10"
/cell-type="ES"
1.1659
/gene="Alre"
1.1659
/gene="Alre"
/codon_start=1
/product="Alre protein"
/protein_id="CAB36909.1"
/db.xref="gi:4456675"
/translation="MAGDGLMLRLRLRLHRTETAYVADSAPFLHLADHDVPEDFK
OETRLKEKECPQAFHALLSWLIRDSQALIDFENLILKPDVLEYSRLHSLIDGPP
KPDVLSNRKGRKPLAGKAVLPRLPRPKRLALDEPRATPATLASKSVSSGSHLK
TKPRKPDGNLESQHLPLGNCIQTMAASVORAVTVASGDPGRGAVEGLLIDQVES
GRSKKLIQVGGEEFTYPNKEFEDPSGNLKNKARSGSSLKPVRAKQAVTLPGRDEKVG
QQCCVPLPLPSEPVQYNQKNEDECAVCHDDELICDDCPRAFHILACSLPQETLPS
GLMCSCLQGRVQOONLSQPEVSRPELPALPEPILVGLRSASEKTSRHELKASSDA
AVTYVNLPLPAPRALPEPSALICPLSSAGNEGRPGAPARASVCGDGEVLICARCA
AFHFRCHPPLPAAAPCPNLKSCSADSTPVRPGAVPRGPAVPLAGLAVGDSS
ASIDPVLHRRDLESLLNHSFDGLIQMLIQSMRPLAETPPSS"
BASE COUNT      345 a      532 c      480 g      302 t
ORIGIN
Query Match      42.9% Score 964.2; DB 10; Length 1659;
Best Local Similarity 76.0%; Pred. No. 1,7e-160;
Matches 1249; Conservative 0; Mismatches 373; Indels 21; Gaps 4;

129 GGACGCGCGCTTACGCGGCGCTTGTGAGCTGCACCGACGAGATCGCGGTGCGTGA 188
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12 GGATGGAATGCTACGCGCTCTGCTGAGGCTGCACCGCACCGAGATCGCGGTGCA 71
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 CAGCGCCTCCCATCGCTGCACGCGCTGGCTGACACGACGAGTGGTCCCGAGACAATT 248
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 CAGTGCTTTCCGTGCTGCAATGCTTGAACGACACGACGAGTGGTCTTGAGGACAATT 131
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 TCAGGAGACGCTTCATCTGAAGAAAGAGAGGCTGCCGCCAGGCTTCACGCGCTCT 308
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 CCAGAGACGCTCGCTGCTGAAGAGAAAGAGGCTGGCCCGACGCGCTTCACGCGCTCT 191
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 GTCTGCGCTCTGCACGAGCATCCACAGCCATCTGTGACTTCTGAGAGGTGCTTTCAA 368
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192 GTCTGCTGCTCTGCACCGGAGCACTGGGGCCATCTGGAATTTCTGAGAGATTCTTTTAA 251
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
369 GGACTACAACTGGAGGCGTATTGGCCGGCTGCAGCCCAATCCGAGAGAGTTCCCGCAACA 428
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 GGACTACAACTGGAGGCGGTACAGCCGCTCATAGCATCTCTGGACGCGCTTCCCAAAA 311
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
429 TGTGACCTCAGCCAGCCCGGAGAGGAGAGGCCCGGCGCTTCCCAAGGCTTTGGT 488
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 TGTGACCTTAAACACAGTCCCGGAAAGGAGAAAGGCCCTTGCTGCCAAGCGCGGT 371
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
489 ACCGCAACCAAGCATCCCAACCAAGAGAGGCGCTCAAGAGAGCTCGAGCTGCGCGCC 548
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 ACTGCCACCAAGACCCCAACCAAGAGAAAGGCACTGGAGAGGCTTCAGACCAACCC 431
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 AGCAGCCCTGATCTCAAGAGGGGACACCGGAGCCAGGCTCTCAATCTGAAGGCCAAGCCCC 608
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
432 AGCAACTCTGGGCTTCAAGAGCGTCTTCAGGCCAGGCTCCCACTGAAAGACTTAAACCCCC 491
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
609 CAAAGAACCGGAGAGCAGCGCAGACGACAGCGGCTTCCACTCGGGAACGGGATTCAGAC 668
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 TAAAGAACCGAGATGGCAACTTGGAGTCAACAGCACTTCCTTGGAAACGGGAATTCAGAC 551
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
669 CATGTCAGTTCAGTCTCAGAGAGCTGTGGCATCTCTCCGGGAGACGTCCCGGAGCCCG 728
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
552 CATGCAAGCTTCTGTCCAGAGAGCTGTGACCGTGGCGTGTGGGATGTTCCAGAAACCCG 611
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
729 AGGGGCGGTGAGAGGGATCTCATCAGACAGGTTTGTGACTCAGAGGCGCTCCAGAAAGNG 788
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
612 AGGGGCGGTGAGAGGGATCTTATCCAGACAGGTGTTTGAAGTCAAGAGAGATCCAGAAAGTG 671
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

Db 372 ACTGCACACCCAGACCCACACAGAGAAAGACCTGGAGGAGCTCGAGCCACCCAC 431
Qy 549 AGCAGCCCTGACTCCAGAGGGGACCCAGCCAGCTTCTCACTGAAGGCCAGCCCG 608
Db 432 AGCAACTCTGCTCAAGAGAGGCTCCAGGCCAGGCTCCACCTGAAAGACTAAGCC 491
Qy 609 CAAGAAGCCGAGAGCAGCAGAGAGAGAGGCTTCCACTCGGGAAGGAGATTCAGAC 668
Db 492 TAAAGACCCAGATGGCAACTTGGAGTACAGACCTTCTTGGAAACGGAATTCCAGAC 551
Qy 669 CATGTCAAGCTTCACTCCAGAGAGCTGTCCATGTCCCGGGAGAGTCCCGGAGCCCG 728
Db 552 CATGGAGACTTCTCTCCAGAGAGCTGTACCTGTGGCTGTGGGATGTTCCAGAACTCG 611
Qy 729 AGGGCCGTGGAGGGAGTCTCTATCCAGAGAGTGTGTAGTACAGCGGCTCCAGAAAGT 788
Db 612 AGGGCCGTGGAGGGAGTCTCTATCCAGAGAGTGTGTAGTACAGAAAGTCCAGAAAGT 671
Qy 789 CATCCAGGTTGGTGGAGTCTCTACCTCCAGCAAGTTCCAGAACTCC--GGCAGTGG 845
Db 672 CATTCAGGTTGGGAGAGTCTTATACACCCAGCAAGTTCCAGAACTCCAGTGGCAATT 731
Qy 846 GAAGAAACAGGCGCCGAGCAGCAGTGGCCGAGAGCTCTGGTGGAGCAAGGAGCCCA 905
Db 732 GAAGAAACAGGCGCCGAGAGTGTACAGCTTAAAGCCAGTGTCCAGAGCCAGAGGAG 791
Qy 906 GGGCGCTGCCCCGGTGGAGGTAGGCTAGGCTGGGCGCAGCAGGCTTCCGCGCC 965
Db 792 GGTACATATACCTGTAGATGAGATGAGAAAGTGGGCGCAGAGTGTGGGTTCCCTCC 851
Qy 966 TCTGGCCCTCCAGAGTACCCAGCTCCACCAAGAAATAGAGAGTGTGCGCTGTG 1025
Db 852 TCCATCCCTCCAGAGTACCCAGCTTAAACCAAGAAAGTGTGCGCTGTG 911
Qy 1026 TCGGAGAGGCGGGGAGCTCTATGCTGAGAGGCTGCGGCTTCCAGCTGGGCTG 1085
Db 912 CCAGCAGGAGGAGTGTAGTCTCTTGTGAGAGGCTGTCCCGGCTTCCAGCTGGCTG 971
Qy 1086 CCTGTCCCTGCGCTCCGAGAGATCCCACTGGAGCTGGAAGTCTCCAGCTGCTCA 1145
Db 972 CCGTCCCTGCGCTCCGAGAGATCCCACTGGAGCTGGAAGTCTGCTGCTGCTCA 1031
Qy 1146 GGCACAGTCCAGAGTGGAGCCCGGCGAGAGAGAGCCCGGCGCCAGAGCCAGCT 1205
Db 1032 GGGCAAGTCCAGAGATCTGTCCAGCTGAGTGTCCAGGCGCCCGGAGTACCTGC 1091
Qy 1206 GGAAGCCCGCTCCCGGCGCTAGGTCGGGCGAGAGAGTAAAGAGTCCAGCTG 1265
Db 1092 AGAGACCCCGATCTGTGGAGTGAAGTCACTTCAAGAAACCAAGGCGCCATCCAG 1151
Qy 1266 GGAACCCCTAGCCGATGAGACAGACTCTTGTCTACAGACACCTGCGGCTCCGCTTC 1325
Db 1152 GGAGCTCAAGAGCCAGCTGTGATCTGTCTGACATATGTAACCTGCTGGCGCCAG 1211
Qy 1326 TGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1385
Db 1212 TGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1265
Qy 1386 TGAAGGTCAGAGAACCTGCTCTGCTGCTGGGCTGAGGAGTGGTGAAGA 1445
Db 1266 TGAAGGTCAGAGAACCTGCTCTGCTGCTGGGCTGAGGAGTGGTGAAGA 1335
Qy 1446 CGTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1505
Db 1336 GGTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1385
Qy 1506 CACCTCCGCGGCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559
Db 1386 CCGCCCGCGGCGCGGAGCAATCTCGGTCGCAATCTCTGCTGCAAGACTGCCAC 1445
Qy 1560 CCGAGCCCTGCTGAGAGGCTGCTGCGCCCGAGCCCGCGGCTGCGGCTGCTGCTG 1619

```

```

Db 1446 GCCAGGACACCGGCGAAGCTGTACCCACCTCTGGGCCCCGTCACAGCCTGGCTGC 1505
Qy 1620 CAA-----GGATGACATCGCATCAAGAGCCCGGCTGTGACAGGATGACCTGATC 1673
Db 1506 CAAGGTAGGGGACACCTCTGTCTACAGACCTGTCTTACATGAGGAGCAGCTGATG 1565
Qy 1674 CATTCTGAGGAGACACACTTGTGATGATCTCAGTGGGCGCATTCAGACATGAGCCG 1733
Db 1566 CTTCTCAATGAGACACTTGTGAGGAGCTCTGACATGCGGCGCATTCAGACATGTGACG 1625
Qy 1734 TCCGGGCGGCGCCCTTCCCTCTCT 1756
Db 1626 CCGCGTGGCGGAGACACCACTT 1648

RESULT 10
AF079536
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1936)
Wang, C. Y., Shi, J. D., Davoodi-Semiromi, A. and She, J. X.
Cloning of Aire, the mouse homologue of the autoimmune regulator
(AIRE) gene responsible for autoimmune polyglandular syndrome type
1 (ASPI)
JOURNAL
Genomics 55 (3), 322-326 (1999)
MEDLINE
99168902
PUBMED
10049587
REFERENCE
2 (bases 1 to 1936)
Wang, C. Y., Shi, J. D., Davoodi-Semiromi, A. and She, J. X.
Direct Submission
Submitted (21-JUL-1998) Pathology, Immunology and Laboratory
Medicine, University of Florida, 1600 SW Archer Road, Gainesville,
FL 32610, USA
FEATURES
source
location/Qualifiers
1..1936
/organism="Mus musculus"
/strain="B6"
/db_xref="taxon:10090"
/tissue_type="thymus"
1..1936
/gene="Aire"
58..1716
/gene="Aire"
/function="transcription factor"
/product="autoimmune regulator"
/protein_id="AAD2044.1"
/db_xref="GI:4426599"
/translation="MAGSDMILRLRLRLHRTETIAVAIDSAFLLHALADHDVPEDEK
OETLRKEGCPQAPFALLSMILTRDSGALIDFWRILFKVNLRYRSRLTILGFP
KDVPLSRKGRKPLAGKAVLPRPTTKKALEPAPATLASVSPGSHLK
TKPRKPKDNGESQPLNGSIQTMAVSQAVVAVSDVPGTGAIVGILIQVEES
GRSKKCIQVGGERTYTPNKFDPSSGMLKKRAGSSSLKPYRAKGAQVTPGPRDEKVG
QOCCVPPLPSLSEPOYNOKNEDBCAVCHDGGELCCGCPRAFLACLSPLOEIPS
QIMRSCCLQGRVQONLSQPSRPELPAETPLIVGRSASEKTRGSRRELKASDA
AVTYVNLPAHPAPLLEPSALCPILLSNGNRGPGASARCSNGDTEYLKCHCA
AAFWHRHLPAPARPGTNLCKKSCSADSTPTPTGPAVPTSGPRAPGLAKVGDDDS
ASHPDVHRRDLESLNDBHSFDGILDMATIOSMSRPLAETPPSS"
BASE COUNT
408 a 609 c 568 g 351 t
ORIGIN
Query Match 42.9% Score 964.2 DB 10; Length 1936;
Best Local Similarity 76.0% Pred. No. 1.6e-160;
Matches 1249; Conservative 0; Mismatches 373; Indels 21; Gaps 4;
Qy 129 GGACGGCGCTTACGCCGCTTGTAGAGCTCACCCGACAGAGATCGCGTGGCGGTGA 188

```



```

Db 69 GATGAAATGCTACCGCCGCTGCTAGGCTGACCGCACCGAGATGCGGTGCCCATAGA 128
QY 189 CAGCCCTTCCACTGCTGACAGCGCTGGCTACCAAGAGAGCTGTCGCCGAGCAAGTT 248
Db 129 CAGTCCCTTCCGCTGCTGATGCTCTAGCCACCAAGAGAGTGTCTCTGAGAGCAAGTT 188
QY 249 TCAGAGAGCGCTTCACTGGAAGAGAGAGAGGAGGCTGCCCCAGGCTTCCAGCCCTCT 308
Db 189 CCAGAGAGCGCTCCCTCTGGAAGAGAGAGAGAGGCTGCCCCAGGCTTCCAGCCCTCT 248
QY 309 GTCCGTGCTGTCGACCCAGAGAGCTCCAGAGCCATCTCTGAGCTTCTGAGAGGTGCTTCAA 368
Db 249 GTCCGTGCTGTCGACCCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
QY 369 GAGCTACAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
Db 309 GAGCTACAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
QY 429 TGTGAGACTTCAAGCCAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
Db 369 TGTGAGACTTCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
QY 489 ACCGCAACCCAGAGCTCCCGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
Db 429 ACTGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 549 AGCAGCCCTGACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
Db 489 AGCAACTCTGCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
QY 609 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
Db 549 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
QY 669 CATGTCAGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
Db 609 CATGTCAGCTTCTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
QY 729 AGGGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 788
Db 669 AGGGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
QY 789 CATCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 845
Db 729 CATCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 788
QY 846 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905
Db 789 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 845
QY 906 GGGGCGTGGCCCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
Db 849 GGTACACTATACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 908
QY 966 TGTGAGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
Db 909 TGTGAGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 968
QY 1026 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1085
Db 969 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1028
QY 1086 CCGTGTCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1145
Db 1029 CCGTGTCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1088
QY 1146 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1205
Db 1089 GGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1148
QY 1206 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265

```

```

Db 1149 AGAGACCCCGATCTCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1208
QY 1266 GGAACCCCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
Db 1209 GGAAGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1268
QY 1326 TGCAGCCCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385
Db 1269 TGCAGAGCTCTCT-----GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1322
QY 1386 TGAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1445
Db 1323 TGAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
QY 1446 GGTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
Db 1383 GGTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442
QY 1506 CAGCTCCCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1559
Db 1443 CGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1502
QY 1560 CCAAGCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1619
Db 1503 GCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
QY 1620 CAA-----GAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1673
Db 1563 CAAAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1622
QY 1674 CCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1733
Db 1623 CCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1682
QY 1734 TCGGAGGCGCCCTCTCTCTCT 1756
Db 1683 CCGCGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705

```

RESULT 11
 AF128115
 LOCUS
 DEFINITION
 Mus musculus autoimmunity regulator (Aire) mRNA, alternatively spliced product 1b, complete cds.
 AF128115
 AF128115.1 GI:7108531
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1933)
 Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
 Expression and alternative splicing of the mouse autoimmunity regulator gene (Aire)
 Unpublished
 2 (bases 1 to 1933)
 Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
 Direct Submission
 Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
 Location/Qualifiers
 1..1933
 /organism="Mus musculus"
 /strain="B6"
 /db_xref="taxon:10090"
 /chromosome="10"
 /map="between D10Mit31 and D10Mit10"
 1..1933
 /gene="Aire"
 58..1713
 /gene="Aire"
 /function="transcription factor"

[illegible]

QY	1206	TCGGAGAGGGGGAGAGCTCATCTGCTGTGACGGCTGCCCTCCTCGGGCCCTTCCACCTGGGCTG	1085
QY	966	TCCTGGCCCTCCCAAGTACACCCCAAGCTTCACACAGAAATGAGGACGAGTGTGCCGTGTG	1025
Db	849	GCTCACTATACCTGTGTGATGATGACAGAAAGTGGGCGAGCATGTGGGTTTCTCCCTT	908
QY	1026	TCGGAGAGGGGGAGAGCTCATCTGCTGTGACGGCTGCCCTCCTCGGGCCCTTCCACCTGGGCTG	1085
Db	966	CCACGAGGAGAGTGTGAGCTCATCTGTTGTGTGACGGCTGTCCCGGGCCCTTCCACCTGGGCTG	1025
QY	1086	CCTTTCCTCCCTCCGCTTCGGGGAGTCCCAAGGGGAGCTTGAGAGTGTCCAGCTGCCGCA	1145
Db	1026	CTTCTCCCACTTGTGAGAGGATCCCAAGTGGCTTGTGAGATGTCTGTGCTGTCCCA	1085
QY	1146	GCGCAACAGTCCAGAGAGTGTGACGCCCCCGGAGAGAGACCCCGCCAGAGAGCCAGT	1205
Db	1086	GGGAGAGTGTCAACAGAAACCTGTCTCCAGCCCTGAGAGTGTCCAGAGCCCGCCGAGCTACCTGC	1145
QY	1206	GAGAGACCCCTCTCCCGCGGGGCTTGAAGTGGCGGAGAGAGAGTAAAGAGTCCACCTGG	1265
Db	1146	AGAGACCCCATCTCTCTGTGGAGCTGAGGTCAAGTTCAGAGAAAAACAGAGGGCCATCAG	1205
QY	1266	GGAACCCCTCCCGGACGTGAGACAGACTCTTGTCTAAACACACTGCCGCTCCGCTTC	1325
Db	1206	GGAGCTCAAAAGCCACAGCTCTATGCTGTGCATATATGTAACCTGCTGGCCCGCACCC	1265
QY	1326	TGCAGCCCGCTCCAGAGGCTGTGACTTCCTCGGCCCTTGACACCCCTACTGTGTGGGTCC	1385
Db	1266	TGCAGCTCTCT-----GCTGAGGCTTGAAGCATGTGCTTACTGTGTGTGGAA	1319
QY	1386	TGAGGCTCAGCACAACCTGGCTCTGTGTGGCTGTGGGCTGTGGAGATGTACGA	1445
Db	1320	TGAGGGGGCGGCGAGTTCACAGCACCAAGCGCGCATCAGTGTGTGGCATGTGCACCGA	1379
QY	1446	CGTGTCCGCTTACTACTGCGCGCTGGCTTCACATGGCGCTGTGCACCTCCACACCG	1505
Db	1380	GGTGTGTGGCTGTGCACTGTGCTGCTTCCATGGCGCTGTGCACCTCCGACGGC	1439
QY	1506	CACCTCCGCGCGGAGGAGGCTTGCCTGTGACAGATCTCTCAGAGACGTG-----AC	1559
Db	1440	CGCGCGCGCGGAGGAGCAATCTCGGTGCAAACTCTGTGTGACATCTCCAC	1499
QY	1560	CCGAGCCCTGTGTGAGAGGGGTGCTGGCCCCCAAGCCCCCGCGCTGGCCCTGGGCTGC	1619
Db	1500	GCGAGGCAACCGGGGGAAGCTGTACCCACCTCTGGGGCCCGCTCCAGCACTGGGCTTGC	1559
QY	1620	CAA-----GATGTACATCCAGCAGACAGAGCGGCTGTGCACAGGAGTACCTGAGATC	1673
Db	1560	CAGAGTAGGGAGCACTGTGCTAGTACAGCACTGTCTTACATTAAGGAGAGCACTTGGAGTTC	1619
QY	1674	CCTTGTGAGCAGACACACCTTGTGATGGATTCCTGTGAGTGGGCAATCCAGAGCATGGCCG	1733
Db	1620	CTTCTTCAATGTAGCACTATATTGAAGGCACTCTGCAAGTGGGCAATCCAGAGCATGTACAG	1679
QY	1734	TCCGGCGGCGCCCTTCCCTCT	1756
Db	1680	CCCGCTGGCGGAGACCACTCT	1702

RESULT 12	
AFI28116	
LOCUS	1924 bp mRNA linear ROD 29-FEB-2000
DEFINITION	Mus musculus autoimmune regulator (Aire) mRNA, alternatively spliced product 1c, complete cds.
ACCESSION	AFI28116
VERSION	AFI28116.1 GI:7108533
KEYWORDS	.
SOURCE	Mus musculus.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 1924)
AUTHORS
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
TITLE
Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)
JOURNAL
Unpublished
2 (bases 1 to 1924)
AUTHORS
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
TITLE
Direct Submission
JOURNAL
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1. 1924
/organism="Mus musculus"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1. 1924
/gene="Aire"
58. 1704
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 1c"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AF36461.1"
/db_xref="GI:7108534"
/translation="MAGSDMLRLRLHRTETAVLDAISAPFLIALADHVDENK
OETLRKEKGGPOAFHALISMLTIRDSGLIDPWRLFLPDYNIERSLSIILDFP
KDYLDNOSRGRPLAGPKAAVLPFRPTRKALREPRATPATLASVSPSHLK
TRPKPFDGLSOLPLNGIOTMAISVQRAVAVAGDVGTGAEGILIQVFES
GPKKCIQVGEFYTNNKEDPSGNLKNKRSGLSPVVAKAGQDEKVKQCG
VPLPLPSEPPOYNKDECAVCHDGEICDGCPRAFHACTSPLODIPSGIMR
CSCCLOGRVQONLSQEVSRPPELPETPLIVGGRASERKGRSRELKASDAVTV
VNLAPPAAPLIPSAICPLISAGNEGRAPASCSUGDBTEVLRCHCAAFH
WCHPPTAARPTNLRCKCSADSTPTPTGFAVTSGRPRAPGLAKVDDSDASHD
PVLRDLSLNEHSFDGLQWAIOISMSPRLAETPPFS"
BASE COUNT 405 a 605 c 567 g 347 t
ORIGIN
Query Match 41.9%; Score 939.8; DB 10; Length 1924;
Best Local Similarity 75.7%; Pred. No. 3.3e-156;
Matches 1243; Conservative 0; Mismatches 367; Indels 33; Gaps 5;
QY 129 GAGCGGGCGCTACGCGGCTTGTGAGGCTGCACCGACGAGATCGCGGTGCGGTGA 188
DB 69 GGATGGAATGCTACGCGGCTGTGCGAGGCTGCACCGACGAGATCGCGGTGCGGTGA 128
QY 189 CAGCGCTTCCCACTGCTGCACGCGCTGCTGACACGAGCTGTCCTCCGAGACAGTT 248
DB 129 CAGTGCCTTCCCGCTGCTGATGCTTACGACGACGAGCTGTCCTCCGAGACAGTT 188
QY 249 TCAGAGAGCGCTCATGTGAAGAAAAGAGGGGTGCCCCAGCGCTTCCAGCGCTTCT 308
DB 189 CAGAGAGAGCGCTCGCTGTGAAGAAAAGAGGGGTGCCCCAGCGCTTCCAGCGCTTCT 248
QY 309 GTCTGCTGCTGCTGACCGACCTCAACGACCATCTGACTTGTGAGAGGGGTGCTTCAA 368
DB 249 GTCTGCTGCTGCTGACCGACCGGACGATCTGATTTCTGAGAGATTTCTTTTAA 308
QY 369 GCACTTCAACCTTGGAGCGCTATGCGCGGCTGCAGCCCATCTTGACAGCTTCCCAAGA 428
DB 309 GGACTACAATCTGGAGCGGTACAGCGCTGATGATCTGTGACGAGCTTCCCAAGAAGA 368
QY 429 TGTGACCTCAGCGACCGCGGAAGGAGAGCGCCCGCGCTCCCAAGCGCTTTGGT 488
DB 369 TGTGACCTTAAACCACTCCCGGAAGGAGAGCGCCCTTGTGCTGCTTCCCAAGCGCGGT 428
QY 489 ACCGCGACCCAGACTCCACCAAGAGAGAGCGCTCAGAGAGCGCTGAGCTGCGCGGCC 548
DB 429 ACTGCGACCCAGACCCCGACCAAGAGAGAGCGCTGAGAGCGCTGAGCGCACCCACAC 488

QY 549 AGCAGCCCTGACTCCCAAGGGGACACCCAGCCAGCGCTTCTCACTGAGAGCCCAAGCCCC 608
DB 489 AGCAACTCTGCGCTCANAAGAGCGCTTCCAGCGCCAGCGCTCCACCTGAAAGATTAAGCCCC 548
QY 609 CAAAGAGCCCGAGAGACGCGCAGACGACGAGCGCTTCTCACTCGGGAAGCGGATTCAGAC 668
DB 549 TAAGAGCCAGATGCGCAACTTGGAGTCACAGACACCTTCTCTTGTGAAACGAAATTCAGAC 608
QY 669 CATGTCAGCTTCACTGTCAGAGAGCGTGGCCATGCTCCCGGGAGCGTCCCGAGACCCG 728
DB 609 CATGTCAGCTTCTGTCAGAGAGCGTGTGACCGTGTGCTTGGGGATGTTCCAGAAACCG 668
QY 729 AGGGCGCTGGAGAGGATCCCTATCCAGAGAGTGTGTGAGTCAGAGCGCTCCAGAAAGTG 788
DB 669 AGGGCGCTGGAGAGGATCCCTATCCAGAGAGTGTGTGAGTCAGAGAGTGTTCAGAAAGTG 728
QY 789 CATTCAGGTTGGTGGGAGTTCATCACTCCAGCAGATTCGAAAGTCC--GGCAGTG 845
DB 729 CATTCAGGTTGGGAGAGTTCATCACTCCAGCAGATTCGAAAGTTCGAAAGTTC 788
QY 846 GAAGAACAAAGGGCCCGAGAGAGAGTGGCCGAGACCTCTGTTGAGCCCAAGAGAGCCCA 905
DB 789 GAAGAACAAAGGGCCCGAGAGAGTGTGACAGCTTAAGCCAGTGTGCGACCAAGAGAGCCCA 848
QY 906 GGGCGCTGCGCCCGGTGGAGAGTGAAGCTTAAGCTGGCCAGACAGGGAGCGTCCGCCCC 965
DB 849 G-----GTTAGNATGAGCAGAAAGTGGCCAGCAGTGTGGGCTTCTCCCT 896
QY 966 TCTGCGCTTCCCGAGTACCCCAAGCTCCACCAAGAAATGAGAGAGTGTGCGGTG 1025
DB 897 TCCATCTCTCCCGAGTACCCCAAGCTTAAACCAAGAAATGAGTGTGCGGTG 956
QY 1026 TCGGGAGGGGGGAGCTATCTGCTGAGAGGCTGCGCTCGGGCTTCCAGCTGCGCTG 1085
DB 957 CCAGCAGGAGGAGTGAAGTCTATGTTGTGAGAGGCTGCGCTCGGGCTTCCAGCTGCGCTG 1016
QY 1086 CCGTCCCTCCCGCTCCGAGAGATCCCAAGTGGAGCTGAGAGTGTCCAGCTGCTGCA 1145
DB 1017 CCGTCCCGACCTCTGAGAGATCCCAAGTGGAGTGTGAGAGTGTGCGCTGCTGCA 1076
QY 1146 GGCACAGTCCAGAGAGTGAAGCCCGGGGAGAGAGAGCCCGGGCCCGAGAGCCACCGT 1205
DB 1077 GGGCAGAGTCCAAACAAACACTGCTCCAGCTGAGAGTGTCCAGAGCCCGGGAGTCACTGCG 1136
QY 1206 GGAAGACCGCGCTCCCGGGGCTTGAAGTGGGGGAGAGAGAGTGAAGAGTCCACTGG 1265
DB 1137 AAGAGCCCGATCTCTGTGGAGTGAAGTGAAGTTCAGAGAAAACAGGGGCGCATCAG 1196
QY 1266 GGAACCCCTAGCGCGCATGAGACAGACTTGTCTAACACACCTGCGCGCTCCGCTTC 1325
DB 1197 GGAAGTCAAGAGCAGCTGTGATGCTGTGATGATGATGATGATGATGATGATGATGATG 1256
QY 1326 TGCAGCCCGCTGCGAGGCTGACTCTGCGCCCTGACACCCCTACTGTGTGGGTCC 1385
DB 1257 TGCAGCTCTCTT-----CTGAGAGCTTACAGACATGTGCTTACTGTGCTGGGAA 1310
QY 1386 TGAAGGTGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1445
DB 1311 TGAAGGTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1370
QY 1446 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1505
DB 1371 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
QY 1506 CACCTCCCGCGCGGAGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559
DB 1431 GCGCCCGCGCGGAGCGGAGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1490
QY 1560 CCGACCGCTGAGAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619
DB 1491 GCGAGGACACCGGCGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1550
QY 1620 CAA-----GGATGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1673


```

|||||
Db 1214 GGGACATGCAAGCTGTGATGAGAGAGTGTGAGAGAGACACCTCTCTCTCACTCTGGAA 1273
QY 1831 GCGGCGCGGCTGGGATCAAGAGAGGAGACAGCCACTCTTGTCTAGTCTGGCTGTAA 1890
Db 1274 GCGGCGCGGCTGGGATCAAGAGAGGAGACAGCCACTCTTGTCTAGTCTGGCTGTAA 1333
QY 1891 CAGCTGTGTCTTGGGAGACACAGCCATCATGCTGTGAAATTAACCTGCGCCAC 1950
Db 1334 CAGCTGTGTCTTGGGAGACACAGCCATCATGCTGTGAAATTAACCTGCGCCAC 1393
QY 1951 TTCTACTGTGAGAGTCCCGGAGAGCTCTCTGCTGTGAGCTACTAAATATA 2010
Db 1394 TTCTACTGTGAGAGTCCCGGAGAGCTCTCTGCTGTGAGCTACTAAATATA 1453
QY 2011 AAATTACGTG 2020
Db 1454 AAATTACGTG 1463

RESULT 15
AF128122 1637 bp mRNA linear ROD 29-FEB-2000
LOCUS AF128122
DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, alternatively
ACCESSION AF128122
VERSION AF128122.1 GI:7108545
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Ruan,Q.C., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
JOURNAL unpublished
2 (bases 1 to 1637)
AUTHORS Ruan,Q.C., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
FEATURES
source location/Qualifiers
1..1637
/organism="Mus musculus"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1637
/gene="Aire"
/cds 58..1287
/function="transcription factor"
/note="alternatively spliced product 3a"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36467.1"
/db_xref="GI:7108546"
/translation="MAGGDMRLRLRLHRTETAVAIIDSNFPLMLADHDVPEDEK
QETLRLEKEGCPQAFHALLSWLITRDSGLIDFNFILFDYULERYSLRSLIDGFP
KDVLDNOSRKGRKPLAGPAVALPRLPRKRLLEPRATPRPATLASKSVSSPSHLK
TKPKKPDGNSLHPLGNGIOTMAASQRAVTVASGVPRGRGAVEGILIOVES
GRSKCIVYGGFFTPNKRKEDPSGNLKNKARSSSLKPVRAAGVATTPGRDQKVG
QDCGVPLPSLSEPOVNOKNEDCAVCHDGEGLICDDCPRAFHLACLSPLDEIPTS
GLMRSCLQAGRVQNLISQPEVSRPELPATPDSPLDILICRLDLSHNRGSRCTH
LWAPSTWACQGRRLC"
BASE COUNT 359 a 511 c 472 g 295 t
ORIGIN
Query Match 32.9%; Score 737.6; DB 10; Length 1637;
Best Local Similarity 79.9%; Pred. No. 1.5e-120;
Matches 982; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

```

```

QY 129 GAGCGCGCCCTACGCCGCTTCTGAGGCTGCACCGCAGAGATCGCGGTGCGCTGGA 188
Db 69 GGATGGAATGCTACGCGCTTCTGCTGAGGCTGCACCGCAGAGATCGCGGTGCGCTGGA 128
QY 189 CAGCGCTTCCACTGCTGTGACGCGGTGACACGACGAGGTGTCCCGGAGCAAGTT 248
Db 129 CAGTGCCTTTCCTGCTGTGATGCTTACGCGACACGAGGTGTCCCTGAGHCAAGTT 188
QY 249 TCAGAGAGCGCTTTCATCTGAAGAAAAGAGGCGTCCCGCAGGCTTTCACGCGCTTCT 308
Db 189 CCAGAGAGCGCTCGTCTGTAAGAGAAAGAGGTGCCCCAGGCTTTCACGCGCTTCT 248
QY 309 GTCTGTGCTGTACCCAGGACTTCACAGCATCTTGTGACTTCTGAGAGGTGCTTTCAA 368
Db 249 GTCTGTGCTGTACCCAGGACTTCACAGCATCTTGTGACTTCTGAGAGGTGCTTTCAA 308
QY 369 GGACTCAACCTGAGCGGCTATGCGCGGCTGACGCGCATCTGAGAGCTTCCCAAGA 428
Db 309 GGACTCAACCTGAGCGGCTATGCGCGGCTGACGCGCATCTGAGAGCTTCCCAAGA 368
QY 429 TGTGACCTCAGCAGCCCGGAGAGGAGAAAGCCCGCGTCCCAAGGCTTGTGT 488
Db 369 TGTGACCTCAGCAGCTTCCCGGAGAGGAGAAAGCCCGCTTGTGTGTGCTTCCCAAGGCGGT 428
QY 489 ACCGCCACCCAGACTCCCAACAGAGAAAGGCTTCAGAAAGGCTTGAGCTGCGCGCC 548
Db 429 ACTGCCACCCAGACCCCGCCCAACAGAGAAAGGCTTGAGAGGCTTGAGAGCCACCCAC 488
QY 549 AGCAGCCCTGACTCCAAAGGGGCGACCGCGCCAGGCTTCACTGAGAGCCCAAGCCCGC 608
Db 489 AGCAGCTTGGCTTCAAAAGGCGCTTCCAGCCAGGCTTCCAGCTGAGAGCTAAGCCCGC 548
QY 609 CAAGAACCGGAGAGCGAGCGCAGAGCAGCAGCCCTTCCACTGCGGAGAGGATTCAGAC 668
Db 549 TAAGAAAGCCAGATGCGCACTTGGAGTGCACAGCAGCTTCTCTTGAAAGGAAATTCAGAC 608
QY 669 CATGTACGTTCACTCAGAGAGCTGTGGCATGTCTCCGGGAGAGCTCCCGGAGCCG 728
Db 609 CATGCGAGCTTGTGTCCAGAGACTGTGACCGGCGCTTGGGAGATGTTCCAGAGCCG 668
QY 729 AGGCGCGTGGAGGAGCTCTTCCAGCAGAGGTTGAGTCAAGGGCGCTCCAAAGATG 788
Db 669 AGGCGCGTGGAGAGGATCTTATCCAGCAGAGTGTGAGTCAAGAGATCCAAAGATG 728
QY 789 CATCAGGTTGGTGGGAGATCTTACACTCCACAGAGTTGGAAGACTTC---GGCAGTGG 845
Db 729 CATCAGGTTGGGAGAGATTTTATACCCCAAGAGTTGGAAGACTTC---GGCAGTGG 788
QY 846 GAAGAACAAGAGCGCCGAGCAGCAGAGTGGCCGGAAGCTTGTGAGGCCAAGGAGCCCA 905
Db 789 GAAGAACAAGAGCGCCGAGTGTGAGCAGCTTAAAGCCAGTGTGAGGCCAAGGAGCCCA 848
QY 906 GGGCGCTGCGCGGAGTGTGAGGCTAGGCTGAGCGCCAGGAGAGGCTTCCCGCC 965
Db 849 GGTCACTATACCTGTGAGAGATGAGCAGAAAGTGGCGCAGCAATGTGGGCTTCTCCCT 908
QY 966 TCTGCGCTTCCCACTGAGCAGCCAGCTCCACAGAGAAATGAGAGAGAGTGTGCTGTG 1025
Db 909 TCCATCCCTCCCACTGAGAGCCAGAGTTAACCAAGAAAGAGAGATGAGTGTGCTGTG 968
QY 1026 TCGGAGAGCGGAGAGCTATCTGCTGTGAGAGGCTGCTCGGCGCTTCCACTGCGCTG 1085
Db 969 CCACAGCAGAGGTGTGAGCTATCTGTGTGAGCAGGTGTGCTCCCGGCTTCCACTGCGCTG 1028
QY 1086 CCGTGCCTTCCCGCGCGGAGATCCCAAGTGGAGAGCTTGGAGAGTGTGCTTCCAGCTGCTGA 1145
Db 1029 CCGTGCCTTCCCGCGCGGAGATCCCAAGTGGAGAGTGTGCTTCCAGCTGCTGA 1088
QY 1146 GGCAAGCTCAGAGAGTGTGAGAGCCCGGAGAGAGAGCCCGCCAGAGCAGCAGCT 1205
Db 1089 GGGCAGAGTCCACAGAGAGTGTGCTCCAGCTGAGAGTGTGCTCCAGAGCCCGGAGCTTACCTGC 1148

```

Oy 1206 GGAGACCCCGCTCCCGGGGCT 1229
||| ||| |||
Db 1149 AGAGACCCCGGACCAATCTCCGCT 1172

Search completed: February 14, 2003, 15:22:42
Job time : 5667.5 secs

FT mutation frameshift and truncated 422 aa protein"
 FT replace(1365,C)
 FT /tag- e
 FT /note- "1-nucleotide insertion"
 FT replace(1085..1097)
 FT /tag- f
 FT /note- "13-nucleotide deletion, results in
 frameshift and truncated 372 aa protein"
 FT replace(1051)
 FT /tag- g
 FT /note- "1-nucleotide deletion"
 FT replace(1309)
 FT /tag- h
 FT /note- "1-nucleotide deletion"
 FT replace(1313)
 FT /tag- i
 FT /note- "1-nucleotide deletion, results in
 frameshift and truncated 478 aa protein"
 FT replace(889,T)
 FT /tag- j
 FT /note- "alters Arg codon to STOP, and truncated
 256 aa protein"
 FT replace(358,T)
 FT /tag- k
 FT replace(374,G)
 FT /tag- l
 FT replace(1052,A)
 FT /tag- m
 FT replace(1094,A)
 FT /tag- n
 FT replace(708,T)
 FT /tag- o
 FT /note- "silent polymorphism"
 FT replace(801,T)
 FT /tag- p
 FT /note- "silent polymorphism"
 FT replace(1317,T)
 FT /tag- q
 FT /note- "silent polymorphism"
 FT replace(1698,T)
 FT /tag- r
 FT /note- "silent polymorphism"
 FT /tag- s
 FT /note- "Alu sequence overlapping with PKL promoter
 starts at this position"
 FT 2191..2195
 FT /tag- t
 FT polyA_signal
 FT 2191..2195
 FT /tag- t
 FT WO9918197-A2.
 FT PN
 FT XX
 FT PD 15-APR-1999.
 FT XX
 FT PF 02-OCT-1998; 98WO-EP06294.
 FT XX
 FT PR 12-NOV-1997; 97EP-0119810.
 PR 02-OCT-1997; 97EP-0117154.
 PR 08-OCT-1997; 97EP-0117398.
 PR XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (MPVU) NAT PUBLIC HEALTH INST.
 PA XX
 PI Aaltonen J, Bjorses P, Horelli-Kuitunen N, Lehnach H,
 PI Palotie A, Peltonen L, Perheentupa J, Yaspö M;
 PI XX
 DR WPI: 1999-287735/24.
 DR P-PSDB; AAY06073.
 DR XX
 PT New polypeptide which co-segregates in mutated form - with
 PT autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 PT Claim 1; Fig 2A; 77Pp; English.
 PT PS
 XX

CC This is the nucleotide sequence of human autoimmune polyglandular
 CC disease type 1 (APGD1 or AIRE) cDNA. It is a composite of 2 clones,
 CC which were obtained from a human adult thymus cDNA library by PCR
 CC (see AAY58609-10), and of a 3'UTR extension PCR product. The encoded
 CC protein (see AAY06073) is a transcription factor or transcription-
 CC associated factor that may associate with vimentin fibres, perhaps
 CC as part of a docking mechanism regulating nuclear translocation.
 CC Aggregates of the mutated protein may prevent formation of vimentin
 CC intermediate filaments. Mutated APGD1 polypeptides co-segregate
 CC with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 CC (APECED). The invention provides vectors and host cells for
 CC preparation of APGD1 polypeptides. APECED-associated mutations
 CC found in the APGD1 gene include: an insertion of 4 nucleotides
 CC (CCTG), normally found at position 1086-1089, at nucleotide
 CC position 1085 or 1090; an insertion of an adenosine at position
 CC 1284; an insertion of a cytosine at position 1365; a deletion of
 CC nucleotides 1085-1097; a deletion of the thymidine at position 1051;
 CC a deletion of the cytosine at position 889; a guanosine to thymidine
 CC thymidine exchange at position 358; an adenosine to guanosine exchange at
 CC position 374; a guanosine to adenosine exchange at position 1052;
 CC and a cytosine to adenosine exchange to position 1094. A claimed
 CC method for testing for carrier status of APECED or for the disease
 CC state involves testing for a mutation in the APGD1 gene or for a
 CC mutated form of the APGD1 polypeptide. APGD1 nucleic acid can be
 CC used in gene therapy.
 CC XX

Sequence 2245 BP; 419 A; 765 C; 697 G; 364 T; 0 other;

Query Match 100.0%; Score 2245; DB 20; Length 2245;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGCCACAGCCGGCGGAGGCCACAGCCCGCGGAGCCGAGCCCAACGAGG 60
 DB 1 CGGGCCACAGCCGGCGGAGGCCACAGCCCGCGGAGCCGAGCCCAACGAGG 60
 QY 61 GCTGCCAGTCTCCGAGGACCCACCGGCTCCGCCACAGCCCGGAGTCCCGCCACACCC 120
 DB 61 GCTGCCAGTCTCCGAGGACCCACCGGCTCCGCCACAGCCCGGAGTCCCGCCACACCC 120
 QY 121 ATGGCAGCAGCGCGCGCTTCTGAGGCTGACCGCAGCGAGATGCGGTG 180
 DB 121 ATGGCAGCAGCGCGCGCTTCTGAGGCTGACCGCAGCGAGATGCGGTG 180
 QY 181 GCGGTGACAGGCGCTTCCACAGTGTGACAGCGGCTGCTACCAACAGCTGTCCCGAG 240
 DB 181 GCGGTGACAGGCGCTTCCACAGTGTGACAGCGGCTGCTACCAACAGCTGTCCCGAG 240
 QY 241 GACAAGTTTCAGAGAGCGCTTCTGAAGGAAGAGAGGCTGCCCGCGGCTTCCAC 300
 DB 241 GACAAGTTTCAGAGAGCGCTTCTGAAGGAAGAGAGGCTGCCCGCGGCTTCCAC 300
 QY 301 GCCCTCTGTCTGCTGCTGACCCAGGACTCCACAGCCATCTGAGAGGTG 360
 DB 301 GCCCTCTGTCTGCTGCTGACCCAGGACTCCACAGCCATCTGAGAGGTG 360
 QY 361 CTGTTCAGGAGTACAACTGAGGCGCTTGTGCGGCTGACGCCATCTGACAGCTTC 420
 DB 361 CTGTTCAGGAGTACAACTGAGGCGCTTGTGCGGCTGACGCCATCTGACAGCTTC 420
 QY 421 CCCAAGATGTGAGCTAGCCAGCCCGGAGAGGAGAGAGCCCGCGGCTGCCAAG 480
 DB 421 CCCAAGATGTGAGCTAGCCAGCCCGGAGAGGAGAGAGCCCGCGGCTGCCAAG 480
 QY 481 GCTTTGGTACGCGCACCCAGACTCCACAGAGAGAGGCTCAGAGAGGTGAGGT 540
 DB 481 GCTTTGGTACGCGCACCCAGACTCCACAGAGAGAGGCTCAGAGAGGTGAGGT 540
 QY 541 GCGGCGCAGAGCCCTTACTTCAAGGGGACCGCCAGCGCTTCACTGAAGGCC 600
 DB 541 GCGGCGCAGAGCCCTTACTTCAAGGGGACCGCCAGCGCTTCACTGAAGGCC 600

601 AAGCCCCCAAGAACCGGAGAGCGCAGAGCAGCAGCGCTTCCACTCGGGAAACGGG 660
 601 AAGCCCCCAAGAACCGGAGAGCGCAGAGCAGCAGCGCTTCCACTCGGGAAACGGG 660
 661 ATTGAGACCATGTACGTTTCAAGTCCAGAGAGCTGTGGCCATGTCTCCGGGAGCTCCG 720
 661 ATTGAGACCATGTACGTTTCAAGTCCAGAGAGCTGTGGCCATGTCTCCGGGAGCTCCG 720
 721 GGAGCCCGAGGGGCGCTGGAGGGGATCTCATCCAGCAGAGCTGTGGTGTAGTCAAGCGCTCC 780
 721 GGAGCCCGAGGGGCGCTGGAGGGGATCTCATCCAGCAGAGCTGTGGTGTAGTCAAGCGCTCC 780
 781 AAGAGTGCATCCAGAGCTGTGGGAGGTTCTACCTCCAGCAGAGTTCGAAGCTCCGGC 840
 781 AAGAGTGCATCCAGAGCTGTGGGAGGTTCTACCTCCAGCAGAGTTCGAAGCTCCGGC 840
 841 AGTGGGAAGAACAGAGCCCGCAGCAGCAGTGGCCGGAAGCTTGTGGTGTAGAGCAAGGGA 900
 841 AGTGGGAAGAACAGAGCCCGCAGCAGCAGTGGCCGGAAGCTTGTGGTGTAGAGCAAGGGA 900
 901 GCCCAGGCGCTGCCCGCGGTGGAGGTAGGCTAGGCTGGCCAGCAGCGGAGCTTCCC 960
 901 GCCCAGGCGCTGCCCGCGGTGGAGGTAGGCTAGGCTGGCCAGCAGCGGAGCTTCCC 960
 961 GCCCCTTGCCCTCCAGTACGCCCGCAGTCCACACAGAAATGAGGAGGAGTGGCC 1020
 961 GCCCCTTGCCCTCCAGTACGCCCGCAGTCCACACAGAAATGAGGAGGAGTGGCC 1020
 1021 GTGTGTGCGGAGCGGGGAGGAGTCTATCTGTGAGCGCTGCCCTTGGGCTTTCACCTG 1080
 1021 GTGTGTGCGGAGCGGGGAGGAGTCTATCTGTGAGCGCTGCCCTTGGGCTTTCACCTG 1080
 1081 GCGTGCCTGCTCCCTCCGCTCCGCGGAGATCCCGAGTGGGAGCTGGAGGTGTCTCACTGC 1140
 1081 GCGTGCCTGCTCCCTCCGCTCCGCGGAGATCCCGAGTGGGAGCTGGAGGTGTCTCACTGC 1140
 1141 CTGAGAGCAAGATCCAGAGAGGTGACGCCCGGCGAGAGAGCGCCGCGCCAGAGACCA 1200
 1141 CTGAGAGCAAGATCCAGAGAGGTGACGCCCGGCGAGAGAGCGCCGCGCCAGAGACCA 1200
 1201 CCCGTGAGACCCGCTCCCGCGGGGCTTAAAGTGGCGGAGAGGTAAGAGGTCCA 1260
 1201 CCCGTGAGACCCGCTCCCGCGGGGCTTAAAGTGGCGGAGAGGTAAGAGGTCCA 1260
 1261 CCTGGGGAACCCCTAGCGCGGCTGACAGACTCTTGTCTAACAAGACCTGCCGGCTCCG 1320
 1261 CCTGGGGAACCCCTAGCGCGGCTGACAGACTCTTGTCTAACAAGACCTGCCGGCTCCG 1320
 1321 CCTTCTGAGCCCGCTGCGAGGCTGAGCTCTGCGCCCTGACACCCCTACTGTGTG 1380
 1321 CCTTCTGAGCCCGCTGCGAGGCTGAGCTCTGCGCCCTGACACCCCTACTGTGTG 1380
 1381 GGTCTGTAGGGTCAAGCAACCTGTGCTGTGCGGGTGTGCGGAGATGAT 1440
 1381 GGTCTGTAGGGTCAAGCAACCTGTGCTGTGCGGGTGTGCGGAGATGAT 1440
 1441 ACGGACGCTGCGGGTGTACACTGCGCGCTGCTTCCACTGTGGGCTGCCACTTCCA 1500
 1441 ACGGACGCTGCGGGTGTACACTGCGCGCTGCTTCCACTGTGGGCTGCCACTTCCA 1500
 1501 GCGGCAACCTCCGCGCGGAGCGGCTGCGCTGCAAGTCTTGTCTCAGAGAGACGTGACC 1560
 1501 GCGGCAACCTCCGCGCGGAGCGGCTGCGCTGCAAGTCTTGTCTCAGAGAGACGTGACC 1560
 1561 CCAAGCCCTGTGAGAGGGGTGTGCTGCGCCCAAGCCCGCGCTGCGCTGCGCTGCC 1620
 1561 CCAAGCCCTGTGAGAGGGGTGTGCTGCGCCCAAGCCCGCGCTGCGCTGCGCTGCC 1620
 1621 AAGGATGACACTGCGCAGTCAAGAGCCCGCTTGTGACAGAGGATGACTGTGAGTCTTCTG 1680
 1621 AAGGATGACACTGCGCAGTCAAGAGCCCGCTTGTGACAGAGGATGACTGTGAGTCTTCTG 1680
 1681 AGCAGACACACTTGTGATGATCTCTGCAAGTGGCCATTCAGAGCATGCGCTCGGCGG 1740

1681 AGCAGACACACTTGTGATGATCTCTGATGAGTGGCCATTCAGAGCATGAGCCCGCTCGGCG 1740
 1741 GCGCCCTTCCCGCTGAGTCCAGATGCGCGGAGATCAGATGATGAGAGTGTGCTG 1800
 1741 GCGCCCTTCCCGCTGAGTCCAGATGCGCGGAGATCAGATGATGAGAGTGTGCTG 1800
 1801 AGAAGACACCTCTTCTCTGATGCTGAGACCGCGCGGTGGATTCAGAGAGGGAGACG 1860
 1801 AGAAGACACCTCTTCTCTGATGCTGAGACCGCGCGGTGGATTCAGAGAGGGAGACG 1860
 1861 GCGCACCCTTGTAGTGTGCGGCTGTAAACAGCTCTGTGTTTCTGGGGACACAGCCAT 1920
 1861 GCGCACCCTTGTAGTGTGCGGCTGTAAACAGCTCTGTGTTTCTGGGGACACAGCCAT 1920
 1921 CATGTGCTGTGAAATTAACCCCTGCGCCACTTCTACTCTGGAAGTCCCGGAGAGCTC 1980
 1921 CATGTGCTGTGAAATTAACCCCTGCGCCACTTCTACTCTGGAAGTCCCGGAGAGCTC 1980
 1981 TCTTGTGCTGTGACCTACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
 1981 TCTTGTGCTGTGACCTACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
 2041 TAATCCAGCTACATGGAGGCTGAGCATGAGTCACTTGAATCTGGAGAGTGGAGGT 2100
 2041 TAATCCAGCTACATGGAGGCTGAGCATGAGTCACTTGAATCTGGAGAGTGGAGGT 2100
 2101 TGCAGTGTGATGATGAGTGGAGCTGAGCATGAGTCACTTGAATCTGGAGAGTGGAGGT 2160
 2101 TGCAGTGTGATGATGAGTGGAGCTGAGCATGAGTCACTTGAATCTGGAGAGTGGAGGT 2160
 2161 CTCAAAAAACAAAAACAAAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220
 2161 CTCAAAAAACAAAAACAAAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220
 2221 TCACTGCAATCAGATCATCATGATTA 2245
 2221 TCACTGCAATCAGATCATCATGATTA 2245

RESULT 2
 AAX26936
 ID AAX26936 standard; DNA: 2036 BP.
 XX
 AC AAX26936;
 XX
 DT 24-JUN-1999 (first entry)
 XX
 DE cDNA encoding a human autoimmune regulator-1 (AIR-1) protein.
 XX
 KW Autoimmune regulator-1; AIR-1; immune maturation; immune response;
 KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
 KW APECED; autoimmune polyglandular syndrome type I; Aps I; ss.
 OS Homo sapiens.
 XX
 FH key
 FT location/Qualifiers
 FT 137..1774
 FT /*tag= a
 FT /product= "AIR-1"
 PN W0915559-A1.
 XX
 PD 01-APR-1999.
 XX
 PF 23-SEP-1998; 98W0-F100749.
 XX
 PR 23-SEP-1997; 97F1-0003762.
 XX
 PA (FIM-) FINNISH IMMUNOTECHNOLOGY LTD.
 XX
 PI Antonarakis S, Heino M, Krohn K, Kudoh J, Lallioti M;
 Peterson P, Scott H, Shimizu N;

Db 1326 GGTGTTGGCGTGTGCACACTGTGCCGTGCTTCCACATGGCGCTGCCACTTCCCGAGGCG 1385
 QY 1506 CACCTCCCGCGCGCGGAGGCGCTCGCTGCAGATCTGCTAGAGAGACGTG-----AC 1559
 Db 1386 CGCGCGCGCGCGCGGAGCAATCTCCGTGCAAAATCTGCTGTGCAGACTCGACCTCCAC 1445
 QY 1560 CCCAGCCCCCTGTGGAGGGGTGCTGGCCCGCCAGCCCCCGCCCTGAGCCCCCTGGGCGCTGC 1619
 Db 1446 GCCAGGACACCGCGCGGAGGCTGTACCCACCTGTGGCGCCCTCCAGCACCTGGCGCTGC 1505
 QY 1620 CAA-----GGATGACATCGCAGTGCAGAGCGCGCTGTGCACAGAGGATGACCTGGAGTC 1673
 Db 1506 CAAGGTAGGGAGGACAGCTGTGCTAGTACAGACCCCTGTCTACATAGGAGAGACTGGAGTC 1565
 QY 1674 CTTTCTGAGCGACACACCTTGTGATGATGATCTGTGCAGTGGCGCATCAGAGCATGGCGCG 1733
 Db 1566 CCTCTCATATAGACACTATTTGACGCAATCTGTGAGTGGCGCATCAGAGCATGTGCAG 1625
 QY 1734 TCGCGGCGCGCGCTTCCCTCT 1756
 Db 1626 CCGCTGCGCGAGACACCACTT 1648

RESULT 5

AAx26938
 ID AAX26938 standard; DNA; 1463 BP.

AC AAX26938;

DT 24-JUN-1999 (first entry)

XX cDNA encoding a human autoimmune regulator-3 (AIR-3) protein.

XX Autoimmune regulator-3; AIR-3; immune maturation; immune response;
 KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
 KM APECED; autoimmune polyglandular syndrome type I; APS I; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 237..1001

FT /*tag= a

FT /product= "AIR-3"

PN MO9915559-A1.

XX 01-APR-1999.

PF 23-SEP-1998; 98WO-F100749.

PR 23-SEP-1997; 97FI-0003762.

XX (FIM-) FINNISH IMMUNOTECHNOLOGY LTD.

PI Antonarakis S, Helno M, Krohn K, Kudoh J, Laiotti M;

PI Peterson P, Scott H, Shimizu N;

DR WPI; 1999-244390/20.

DR P-PSDB; AAY01714.

XX Autoimmune regulator 1 (AIR1) DNA sequence

PS Claim 4; Page 31-32; 59pp; English.

XX The present sequence encodes an autoimmune regulator-3 (AIR-3)
 CC protein. The AIR polypeptides and polynucleotides can be used
 CC in methods for the diagnosis and treatment of diseases related
 CC to immune maturation and regulation of immune response towards
 CC self and nonself. They can be used particularly in the diagnosis
 CC and treatment of autoimmune polyendocrinopathy candidiasis
 CC ectodermal dystrophy (APECED) (also known as autoimmune
 CC polyglandular syndrome type I (APS I)).

SO Sequence 1463 BP; 245 A; 456 C; 486 G; 276 T; 0 other;

Query Match 37.9%; Score 851.2; DB 20; Length 1463;
 Best Local Similarity 91.7%; Pred. No. 3,7e-157;
 Matches 945; Conservative 0; Mismatches 3; Indels 82; Gaps 1;

QY 991 CTCACCAAGAAATGAGAGAGAGTGTCCGTGTGTCGGAGAGCGGGGAGCTCATCTGC 1050
 Db 516 CTGTGCGAAGAAATGAGAGAGAGTGTCCGTGTGTCGGAGAGCGGGGAGCTCATCTGC 575
 QY 1051 TGTAGAGGCTGCCCTCGGAGCTTCACCTGAGCTGCTGTCCTGCTCGCTCGGAGATC 1110
 Db 576 TGTAGAGGCTGCCCTCGGAGCTTCACCTGAGCTGCTGTCCTGCTCGGAGATC 635
 QY 1111 CCGAGTGGACCTGTGAGAGTGTCCAGTGTCCGTGAGGAGCAAGTCCAGAGAGTGCAGCC 1170
 Db 636 CCGAGTGGACCTGTGAGAGTGTCCAGTGTCCGTGAGGAGCAAGTCCAGAGAGTGCAGCC 695
 QY 1171 CGGCGAGAGAGCCCGCGCGCCAGAGGACACCCGTGAGACCCCGCTCCCGGGGCTT 1230
 Db 696 CGGCGAGAGAGCCCGCGCGCCAGAGGACACCCGTGAGAGACCCCGCTCCCGGGGCTT 755
 QY 1231 AGGTGCGGGAGAGAGGTAAAGAGTCCACCTGGGGAAACCCCTAGCGGATGAGACAG 1290
 Db 756 AGGTGCGGGAGAGG----- 771
 QY 1291 ACTCTGTGTACAAAGACACTGTGCGGCTCCGCTTCTGAGCGCCGCTGCCAGGCTGAG 1350
 Db 772 -----AGCCCGCTGCCAGGCTGAG 793
 QY 1351 TCTCGGCGCTGCACCCCTACTGTGTGTGGTCTGTAGGCTCAGACAGACCTGTCTCT 1410
 Db 794 TCTCGGCGCTGCACCCCTACTGTGTGTGGTCTGTAGGCTCAGACAGACCTGTCTCT 853
 QY 1411 GGTGCGGCTGTGGGGGTGTGGGAGTGTACGAGAGTGTGCGGTGTACTACTGTGCGCC 1470
 Db 854 GGTGCGGCTGTGGGGGTGTGGGAGTGTACGAGAGTGTGCGGTGTACTACTGTGCGCC 913
 QY 1471 GCTGCTTTCACATGCGCTGCGCTGCTCCAGCCGCGACCTCCGCGCGGAGCGGCTGTG 1530
 Db 914 GCTGCTTTCACATGCGCTGCGCTGCTCCAGCCGCGACCTCCGCGCGGAGCGGCTGTG 973
 QY 1531 CGCTGAGATCTGTCTGTAGAGAGCTGACCCAGCCCTGTGTGAGAGGGGTCTGTGCGCC 1590
 Db 974 CGCTGAGATCTGTCTGTAGAGAGCTGACCCAGCCCTGTGTGAGAGGGGTCTGTGCGCC 1033
 QY 1591 AGCCCGCGCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1650
 Db 1034 AGCCCGCGCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1093
 QY 1651 CTGCACAGGATGACCTGTGAGTGTCTGTGAGAGAGTGTGAGAGAGTGTGAGAGTGTGAG 1710
 Db 1094 CTGCACAGGATGACCTGTGAGTGTCTGTGAGAGAGTGTGAGAGAGTGTGAGAGTGTGAG 1153
 QY 1711 TGGGCGATCCAGAGATGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1770
 Db 1154 TGGGCGATCCAGAGATGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1213
 QY 1771 GGGACATGACGCTGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1830
 Db 1214 GGGACATGACGCTGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1273
 QY 1831 GCGGCGCGCTGTGAGATCAAGAGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGTAA 1890
 Db 1274 GCGGCGCGCTGTGAGATCAAGAGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGTAA 1333
 QY 1891 CAGCTGTGTGTTTGTGGGAGACACAGGATGTGCTGTGAGAAATTAACCTGCGCCAC 1950
 Db 1334 CAGCTGTGTGTTTGTGGGAGACACAGGATGTGCTGTGAGAAATTAACCTGCGCCAC 1393
 QY 1951 TTCTTACTGTGAAGTCCCGGGAGGCTCTCTGCTGTGTGACTACTTAAATATTA 2010
 Db 1394 TTCTTACTGTGAAGTCCCGGGAGGCTCTCTGCTGTGTGACTACTTAAATATTA 1453

23-AUG-2000; 2000US-0227009.

[illegible]

XX DE Human genomic DNA for PHIP/NDRP.
XX XX
XX PHIP; Pleckstrin homology domain-interacting protein; NDRP; ds;
XX neuronal differentiation-related protein; insulin receptor substrate;
XX IRS; signal transducer and activator of transcription; STAT;
XX transgenic animal; diabetes mellitus type 2; hyperglycaemia;
XX myotonic muscular dystrophy; acanthosis; nigricans; retinopathy;
XX nephropathy; arteriosclerosis; peripheral arterial disease; cancer;
XX adenocarcinoma; leukaemia; breast cancer; prostate cancer; colon cancer;
XX ovarian cancer; autoimmune disease; inflammation; immunodeficiency.
XX
OS Homo sapiens.
XX
XX WO200185785-A2.
XX
XX PD 15-NOV-2001.
XX
XX PF 10-MAY-2001; 2001WO-CA00673.
XX
XX PR 11-MAY-2000; 2000US-203561P.
XX
XX PA (ROZA/) ROZAKIS-ADCOCK M.
XX PA (FARH/) FARHANG-FALLAH J.
XX PA (CHEN/) CHENG A.
XX
XX PI Rozakis-Adcock M, Farhang-Fallah J, Cheng A;
XX
XX WPI; 2002-041586/05.
XX
XX PT Novel Pleckstrin homology domain interacting protein recruiting
XX proteins of Insulin receptor substrate family, and signal transducer
XX and activator of transcription factors to their receptors, useful to
XX treat diabetes
XX
PS Disclosure: Page 99-133; 139pp; English.
XX
XX The invention relates to an isolated pleckstrin homology domain
XX interacting protein (PHIP) that recruits proteins of the insulin
XX receptor substrate (IRS) family, and signal transducer and activator of
XX transcription (STAT) transcription factors, to receptors that interact
XX with and phosphorylate the proteins and STAT transcription factors.
XX the nucleic acid encoding PHIP (nPHIP), a nucleic acid which binds to
XX nPHIP or regions of it, analogues, fragments or allelic variants of PHIP
XX or nPHIP, a nucleic acid sequence having substantial identity or
XX sequence similarity with a nucleic acid sequence fully defined
XX human neuronal differentiation-related protein (NDRP) nucleic acid
XX sequence or its exons as given in the specification, expression
XX vectors and host cells expressing the nucleic acids, anti-PHP antibodies,
XX and a transgenic animal not already expressing PHIP. The nucleic
XX acids, proteins and antibodies are useful for diagnosis and treatment of
XX a condition associated with an insulin receptor (e.g. diabetes mellitus
XX type 2, hyperglycaemia, myotonic muscular dystrophy, acanthosis,
XX nigricans, retinopathy, nephropathy, arteriosclerosis, peripheral
XX arterial disease) or cancer (e.g. adenocarcinoma, leukaemia, breast
XX cancer, prostate cancer, colon cancer, ovarian cancer and many others
XX given in the specification), autoimmune disease, inflammation and
XX immunodeficiency. The protein is also useful for discovering
XX or testing compounds which may be either enhancers or inhibitors of PHIP
XX function. The present sequence is genomic DNA encoding PHIP and NDRP.
XX
SQ Sequence 140036 BP; 48500 A; 25843 C; 22499 G; 43194 T; 0 other;
Query Match 6.8%; Score 152; DB 24; Length 140036;
Best Local Similarity 79.9%; Pred. No. 2,5e-20;
Matches 179; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

DB 1994 AGGGGGCTAGAGCAGAGAAATCTTTGAACCCAGGAGGGGTTGCAGTGCAGTGA 2053
QY 2116 TTGCGCCACTGCAGTCACTGCTGTCGCGAAGAGTGAGACTCGTCTCAAAAACAAACA 2175
DB 2054 TCACGACATGACATCCAGCTCGGGCGACACAGAGGAGAGACTCCATCTCAAAAACAAACA 2113
QY 2176 AAAAACACATATACATTAATTTATCATCTGACACACTTTTACG 2219
DB 2114 AAACAAACAAACAAACACATGCTTAAGAGACTTACCAAGGTCAG 2157
RESULT 11
ABA20762/C
ID ABA20762 standard; DNA; 26427 BP.
XX
XX AC ABA20762;
XX
XX DT 23-JAN-2002 (first entry)
XX
XX DE Human nervous system related polynucleotide SEQ ID NO 13093.
XX
XX KW Human; nootropic; neuroprotective; cytosolic; dermatological; virocidic;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnereary;
KW antiparkinsonian; antischilling; antianemic; antiarthritic; cancer;
KW antineumatic; hepatotropic; cerebrotrophic; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; neurotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX PN WO200159063-A2.
XX
XX PD 16-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01334.
XX
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226688.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.


```

PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 01-DEC-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251160.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251868.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0251997.
PR 05-JAN-2001: 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX Disclosure: SEQ ID NO 13094; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI4678-ABBI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 26427 BP; 5634 A; 7262 C; 7074 G; 6457 T; 0 other;
SQ
Query Match 6.8%; Score 151.6; DB 22; Length 26427;
Best Local Similarity 82.3%; Pred. No. 2.4e-20;
Matches 186; Conservative 0; Mismatches 39; Indels 1; Gaps 1.
QY 1996 CTACTAATAATATAAATTAGCTGGTGTGGTGGTGGCTGGAATCCAGCTACT 2055
Db 4490 CTACTAATAATAAATAAATTAGCTGGTGTGTGGTGGTGGCTGTAATCTGACTACT 4431
OY 2056 GGGAGCCTGAGGCGATGAGAACTGCTTGAACTCGGAGGTGAGGCTTGCAGTAGTGAGA 2115
Db 4430 GGGAGGCTGAGGCGAGGAGAAATTGCTTGAACCCGGAGGTGAGGCTTGCAGTAGTGAGA 4371
OY 2116 TTGGCCACCTGCTCCGCTGTGGTGGCAGAGAGTGAGAGCTTCGTTCAAAACAAACA 2175
Db 4431 TTGGCCACCTGCTCCGCTGTGGTGGCAGAGAGTGAGAGCTTCGTTCAAAACAAACA 2175

```

Db	4370	TTTCACCACTGCAGCTCCAGCGTGGGCAAC- AGAGCCGAGACTCCGTCTCAAAAAAAAAAAAAA 4312
Qy	2176	AAAAAACACATTAACAATAATTATCATCTCGACCACTTTTCAGTT 2221 - - - - - - - - - - - - AAAAAAAAAAAAAAAAAGATTACTTTAAATTCCTACCTCTTCTACTTT 4266
Db	4311	AAAAAAAAAAAAAAAAAGATTACTTTAAATTCCTACCTCTTCTACTTT 4266
	RESULT 13	
ID	ABO8186	
XX	ABO8186 standard; CDNA; 169739 BP.	
AC		
XX	ABO8186;	
DT	18-SEP-2002 (first entry)	
DE	Human osteoblast differentiation related cdna seq ID NO 93.	
XX		
KM	Human; osteoblast; stem cell differentiation; bone tissue deposition; osteoporosis; osteopathic; ss.	
XX		
OS	Homo sapiens.	
PX		
PN	WO200250301-A2.	
PD		
XX	27-JUN-2002.	
PF	18-DEC-2001; 2001WO-US48276.	
XX		
PR	18-DEC-2000; 2000US-255882P.	
PR	24-APR-2001; 2001US-285691P.	
XX		
PA	(GENE-) GENE LOGIC INC. (PROC) PROCTER & GAMBLE CO.	
XA		
PI	Jl D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A; Mertz L;	
PI		
DR	WPI; 2002-557663/59.	
XX		
PT	Use of genes and their expression profiles associated with osteoblast differentiation for screening modulators bone formation, for diagnosing or treating e.g. osteoporosis, or as markers for the differentiation process _	
PT		
XX		
CS	Claim 1; SEQ ID NO 93; 78bp + Sequence Listing; English.	
XX		
CC	The invention relates to genes and their expression profiles are used for:	
CC	(a) screening modulators of precursor stem cell differentiation into osteoblasts, or bone tissue deposition;	
CC	(b) diagnosing abnormal deposition of bone tissue, abnormal rate of osteoblast formation or osteoporosis; or	
CC	(c) treating or monitoring treatment of the conditions cited in (b), or monitoring the progression of bone tissue deposition.	
CC	Specific conditions include postmenopausal osteoporosis, glucocorticoid osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-induced abnormalities in bone formation or bone loss, conditions that involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis), skeletal disease linked to breast cancer, mastocytosis, Fancconi syndrome or fibrous dysplasia. The present sequence is that of an osteoblast differentiation associated cDNA marker of the invention.	
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
CC		
XX		
SQ	Sequence 169739 BP; 49809 A; 35660 C; 35715 G; 48555 T; 0 other;	
Query Match	6.8%; Score 151.6; DB 24; Length 169739;	
Best Local Similarity	85.4%; Pred.No. 3.1e-20;	
Matches 169; Conservative 0; Mismatches 29; Indels 0; Gaps 0;		
1996 CTACHTAAAATTAATAATTAGCTGGTGCTGGTGCTGGCTGTAATCCAGGTACAT 2055		

Db 213 ATTCTGCCAA 204

RESULT 2

US-09-764-864-1715
; Sequence 1715, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1715
; LENGTH: 12493
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1715

Query Match
Best Local Similarity 84.5%; Score 150.4; DB 10; Length 12493;
Pred. No. 3.5e-21;
Matches 169; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1996 CTCTATAAATATATAATAGCTGGTGTGTGGTGGCTGTAAATCCAGCTACAT 2055
|||||
Db 11407 CTCTATAAATATATAAATTAGCGGGGTGTGTGGCTGTAAATCCAGCTACAT 11466
|||||
QY 2056 GGGAGCCTGAGGAGTGAATCACTTGAATCGGGAGGTGAGTGCAGTGCAGAGA 2115
|||||
Db 11467 AGGAGGCTGAGGAGGAGGAGTGAATCACTTGAATCGGGAGGTGAGTGCAGAGA 11526
|||||
QY 2116 TTGCGGCACCTGACCTCCAGTCTGTGTGGCAGAGTGAAGTCCGCTCAAAAACAAACA 2175
|||||
Db 11527 TCGCACCATTGACACTCCAGCTGAGCATTAAGCAGAACTCCGCTCAAAAACAAACA 11586
|||||
QY 2176 AAAAAACACATTAACATATA 2195
|||||
Db 11587 AAAAAAAGATAGTAA 11606
|||||

RESULT 3

US-09-989-920-58
; Sequence 58, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sel-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 3535
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-989-920-58

Query Match
Best Local Similarity 6.7%; Score 150.2; DB 9; Length 3535;
Pred. No. 3e-21;
Matches 176; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1996 CTACTAAATATATAAATTTAGCTGGTGTGTGGTGGCTGTAAATCCAGCTACAT 2055

Db 1307 CTACTAAATATATAAATTTAGCTGGTGTGTGGTGGCTGTAAATCCAGCTACAT 1366
|||||

QY 2056 GGGAGCCTGAGGAGTGAATCACTTGAATCGGGAGGTGAGTGCAGTGCAGAGA 2115
|||||

Db 1367 GGGAGGCTGAGGAGGAGTGAATCACTTGAATCGGGAGGTGAGTGCAGTGCAGAGA 1426
|||||

QY 2116 TTGCGGCACCTGACCTCCAGTCTGTGTGGCAGAGTGAAGTCCGCTCAAAAACAAACA 2175
|||||

Db 1427 TCGTGCACCTGACCTCCAGCTGTGGGTGACAGAGTGAAGTCCGCTCAAAAACAAACA 1486
|||||

QY 2176 AAAAAACACATTAACATTTATCATCTGACCATTT 2214
|||||
Db 1487 AAAAAAAGTAAACAAATTAAGCTATGCCCATTT 1525
|||||

RESULT 4

US-09-908-711-160/c
; Sequence 160, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882

```

? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01347
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 896
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01307
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 864
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01341
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 856
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01336
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 868
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01312
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 60/179, 065
? PRIOR FILING DATE: 2000-01-31
? PRIOR APPLICATION NUMBER: 60/180, 628
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: 60/209, 467
? PRIOR FILING DATE: 2000-06-07
? NUMBER OF SEQ ID NOS: 167
? SOFTWARE: Patentln Ver. 2.0
? SEQ ID NO 160
? LENGTH: 1186
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (342)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (343)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (344)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (345)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (346)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (347)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (348)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (349)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (350)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (351)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (352)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (353)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (354)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (355)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (356)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (357)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (358)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (359)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (360)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (361)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (362)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (363)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (364)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (365)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (366)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (367)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (368)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (369)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (370)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (371)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (372)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (373)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (374)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (375)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (376)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (377)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (378)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (379)
? OTHER INFORMATION: n equals a,t,g, or c

```

```

1 LOCATION: 6058..6076
2 OTHER INFORMATION: 99-16073..rp
3 NAME/KEY: primer_bind
4 LOCATION: 6493..6512
5 OTHER INFORMATION: 99-16073..pu complement
6 NAME/KEY: primer_bind
7 LOCATION: 9661..9678
8 OTHER INFORMATION: 99-16074..rp
9 NAME/KEY: primer_bind
10 LOCATION: 10119..10136
11 OTHER INFORMATION: 99-16074..pu complement
12 NAME/KEY: primer_bind
13 LOCATION: 14754..14774
14 OTHER INFORMATION: 99-13817..pu
15 NAME/KEY: primer_bind
16 LOCATION: 15183..15203
17 OTHER INFORMATION: 99-13817..rp complement
18 NAME/KEY: primer_bind
19 LOCATION: 20137..20155
20 OTHER INFORMATION: 99-16066..pu
21 NAME/KEY: primer_bind
22 LOCATION: 20569..20588
23 OTHER INFORMATION: 99-16066..rp complement
24 NAME/KEY: misc_binding
25 LOCATION: 3239..3285
26 OTHER INFORMATION: 99-16063-218..probe
27 NAME/KEY: misc_binding
28 LOCATION: 6208..6254
29 OTHER INFORMATION: 99-16073-282..probe
30 NAME/KEY: misc_binding
31 LOCATION: 9848..9894
32 OTHER INFORMATION: 99-16074-266..probe
33 NAME/KEY: misc_binding
34 LOCATION: 14945..14991
35 OTHER INFORMATION: 99-13817-215..probe
36 NAME/KEY: misc_binding
37 LOCATION: 20236..20282
38 OTHER INFORMATION: 99-16066-123..probe
39 NAME/KEY: primer_bind
40 LOCATION: 3243..3261
41 OTHER INFORMATION: 99-16063-218..mis
42 NAME/KEY: primer_bind
43 LOCATION: 3263..3281
44 OTHER INFORMATION: 99-16063-218..mis complement
45 NAME/KEY: primer_bind
46 LOCATION: 6212..6230
47 OTHER INFORMATION: 99-16073-282..mis
48 NAME/KEY: primer_bind
49 LOCATION: 6232..6250
50 OTHER INFORMATION: 99-16074-266..mis complement
51 NAME/KEY: primer_bind
52 LOCATION: 9852..9870
53 OTHER INFORMATION: 99-16074-266..mis
54 NAME/KEY: primer_bind
55 LOCATION: 9872..9890
56 OTHER INFORMATION: 99-16074-266..mis complement
57 NAME/KEY: primer_bind
58 LOCATION: 14949..14967
59 OTHER INFORMATION: 99-13817-215..mis
60 NAME/KEY: primer_bind
61 LOCATION: 14969..14987
62 OTHER INFORMATION: 99-13817-215..mis complement
63 NAME/KEY: primer_bind
64 LOCATION: 20240..20258
65 OTHER INFORMATION: 99-16066-123..mis
66 NAME/KEY: primer_bind
67 LOCATION: 20260..20278
68 OTHER INFORMATION: 99-16066-123..mis complement
69 NAME/KEY: misc.feature
70 LOCATION: 285.982,2264,2273,2823,11227,11232,13663,13681,13842..13843
71 OTHER INFORMATION: n=a, y, c or t
72 GS-09-416-384A-2

```



```

RESULT 6
US-09-969-708-79/c
: Sequence 79, Application US/09969708
: Patent No. US20020102532A1
: GENERAL INFORMATION:
: APPLICANT: Augustus, Meena
: TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature

```

Query Match	6.7%	Score	149.4	DB	10	Length	145831
Best Local Similarity	84.4%	Pred. No.	9e-21				
Matches	168	Conservative	0	Mismatches	31	Indels	0
				Gaps			0

RESULT 7
US-09-954-456-2116/c
; Sequence 2116, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul

Query Match	6.7%	Score 149.4	DB 10	Length 145831
Best Local Similarity	84.4%	Pred. No. 9e-21		
Matches 168; Conservative	0	Mismatches 31	Indels 0	Gaps 0

Qy	18162	18162	18144
AAAAACACATACATTA	AAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAA	18144

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-802

```


Db 1345412 CTACTAAATAATATGAAAAATACCTGGTGTGGCGGCGCTGTATCCAGCTACTC 1345353
QY 2056 GGGAGCGCTGAGCATGAGATCTGTAACCTGGGAGGCTGAGTTCAGTGAAGTACAGA 2115
Db 1345352 AGGAGCGCTGAGCGAGCAAGATACCTTGAACCGGGAGGAGCAAGGTTGAGTGAAGTGAAGA 1345293
QY 2116 TTGGCCCACTGCACTGCTGCTGGTGGCGAGAGTGAAGTCCGCTCAAAAACAAACA 2175
Db 1345292 TCGTGCCATTGCTACCTCCAGCGCTGGGAGACAGCAAGCAAGCTGTCTCAAAAAA 1345233
QY 2176 AAAAACCACATA 2188
Db 1345232 AAAAATAAAGA 1345220

RESULT 14
US-09-795-686-1/c
; Sequence 1, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Guicher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: r-g or a
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: y=c/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m-a or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k-g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s-g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w-a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b-g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d-a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h-a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v-a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: n-a or g or c or t/u
US-09-795-686-1

Query Match 6.5%; Score 146.6; DB 10; Length 1503841;
Best Local Similarity 85.0%; Pred. No. 5.1e-20;

Matches 164; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1996 CTACTAAATAATATGAAAAATACCTGGTGTGGCGGCTGTATCCAGCTACTC 2055
Db 1345412 CTACTAAATAATATGAAAAATACCTGGTGTGGCGGCTGTATCCAGCTACTC 1345353
QY 2056 GGGAGCGCTGAGCATGAGATCTGTAACCTGGGAGGCTGAGTTCAGTGAAGTACAGA 2115
Db 1345352 AGGAGCGCTGAGCGAGCAAGATACCTTGAACCGGGAGGAGCAAGGTTGAGTGAAGTGAAGA 1345293
QY 2116 TTGGCCCACTGCACTGCTGCTGGTGGCGAGAGTGAAGTCCGCTCAAAAACAAACA 2175
Db 1345292 TCGTGCCATTGCTACCTCCAGCGCTGGGAGACAGCAAGCAAGCTGTCTCAAAAAA 1345233
QY 2176 AAAAACCACATA 2188
Db 1345232 AAAAATAAAGA 1345220

RESULT 15
US-09-818-512-3
; Sequence 3, Application US/09818512
; Patent No. US20020142416A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: C1001192
; CURRENT APPLICATION NUMBER: US/09/818,512
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 116592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(116592)
; OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

Query Match 6.5%; Score 146.4; DB 10; Length 116592;
Best Local Similarity 84.2%; Pred. No. 3.4e-20;
Matches 165; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 1996 CTACTAAATAATATGAAAAATACCTGGTGTGGCGGCTGTATCCAGCTACTC 2055
Db 139599 CTACTAAATAATATGAAAAATACCTGGTGTGGCGGCTGTATCCAGCTACTC 39658
QY 2056 GGGAGCGCTGAGCATGAGATCTGTAACCTGGGAGGCTGAGTTCAGTGAAGTACAGA 2115
Db 139659 AGGAGCGCTGAGCGAGCAAGATACCTTGAACCGGGAGGAGCAAGGTTGAGTGAAGTGAAGA 1345293
QY 2116 TTGGCCCACTGCACTGCTGCTGGTGGCGAGAGTGAAGTCCGCTCAAAAACAAACA 2175
Db 139719 TTGTGCCACTGCACTGCACTGCGGAGCAAGATGTGACTCATCTCAAAAAA 39778
QY 2176 AAAAACCACATA 2191
Db 139779 AAAAATAAAGA 39794

Search completed: February 14, 2003, 19:51:07
Job time : 2394.54 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:51:42 ; Search time 21 Seconds

(without alignments)
2494.917 Million cell updates/sec

Title: US-09-509-595B-2

Perfect score: 2902

Sequence: 1 MATDAAIRLLRLRHREIAV.....DGLQWALQSMARPAAPPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273.5	9.4	530	2 I38558	Mi-2 autoantigen 2
2	221.5	7.6	3938	2 T42761	Basoon protein -
3	219.5	7.6	1829	2 T34239	hypothetical prote
4	219	7.5	371	2 B49515	phosphoprotein 75
5	217	7.5	1051	2 S55259	Tf1 protein - mou
6	206	7.1	1787	2 T20160	hypothetical prote
7	194.5	6.7	4957	2 T03455	ALR protein - huma
8	194.5	6.4	5262	2 T03455	ALR protein - huma
9	186.5	6.4	607	2 S58151	hypothetical prote
10	179.5	6.2	404	2 S62475	PHD finger protein
11	178.5	6.2	3942	2 T42730	Basoon protein -
12	171	5.9	825	2 T02518	hypothetical prote
13	168.5	5.8	660	1 Q0BE3	BHLF1 protein - hu
14	168	5.8	1237	2 T45070	protein kinase hom
15	161	5.5	684	2 S52835	hypothetical prote
16	160	5.5	350	2 G01950	hypothetical prote
17	159	5.5	958	2 H84783	probable PHD-type
18	157.5	5.4	1201	2 T29329	hypothetical prote
19	155	5.3	839	2 T45908	hypothetical prote
20	154	5.3	1146	2 A38587	collagen, cornea-s
21	153	5.3	1006	2 T42731	atrophin-1 related
22	152.5	5.3	813	2 E85135	hypothetical prote
23	152.5	5.3	839	2 F75518	hypothetical prote
24	152	5.2	810	2 G01252	small GTP binding
25	152	5.2	1049	1 CGB075	collagen alpha 1(I
26	152	5.2	1204	2 T19918	hypothetical prote
27	152	5.2	1736	2 T00391	hypothetical prote
28	151.5	5.2	728	2 S57142	hypothetical prote
29	151.5	5.2	1479	2 T17401	transcription regu

30	151.5	5.2	1532	2 A61262	collagen alpha 1(X
31	150.5	5.2	801	2 T29018	hypothetical prote
32	150	5.2	2274	2 T30258	adenomatous polyo
33	149.5	5.2	1151	2 T18535	high molecular mas
34	149.5	5.2	1320	2 JC5630	TCOF1 protein - mo
35	149	5.1	1207	2 T00378	KIAA0641 protein -
36	149	5.1	2187	2 T30826	nascent polypeptid
37	148.5	5.1	283	2 T11644	hypothetical prote
38	148.5	5.1	1400	2 T31555	hypothetical prote
39	148	5.1	640	2 T08179	LRG5 protein - chl
40	147	5.1	574	2 T43556	Wiskott-Aldrich sy
41	147	5.1	1093	2 T38533	AF17 protein - hum
42	146	5.0	574	2 T38519	Wiskott-Aldrich sy
43	144.5	5.0	1214	2 JC2069	zinc-finger protei
44	144	5.0	455	2 A87913	protein B0205.10 f
45	144	5.0	649	2 T46500	hypothetical prote

ALIGNMENTS

RESULT 1

138558

Mi-2 autoantigen 240 kDa protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 05-Nov-1999

C:Accession: I38558

R:Ge, O.; Nilasena, D.S.; O'Brien, C.A.; Frank, M.B.; Taroff, I.N.

J. Clin. Invest. 96, 1730-1737, 1995

A:Title: Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2

A:Reference number: I38558; MUID:96013633; PMID:7560064

A:Accession: I38558

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-530 <RES>

A:Cross-references: EMBL:008379; NID:9761717; PIDN:AAC50228.1; PID:9761718

Query Match

Best Local Similarity 23.6%; Pred. No. 7.1e-09;

Matches 102; Conservative 43; Mismatches 160; Indels 127; Gaps 15;

QY	116	PAVPKALVPPRL-----PIRKASEARAAAPALPRGTASPSQILAKPPKPESSA	170
DB	116	PSGPALPPADIDPPPIRAKTKGK-----GPHKRRKSPRPD-----	159
QY	171	EQRLPLNGNIGTMSASVQRAVAMSSGDVPRGARVESILIQVFESGSKCIQVGEEF	230
DB	160	-----GRKKLRGKKMAPLKIILGLGKR-----KKGSS--DEGPEP	195
QY	231	YTSKFEFSSGSKKRASSSQPKFLYRAKGAOGAAPPGGGEAR-LGQGSVPAPLALPSDP	289
DB	196	EAEESDLSSSVHSASGRPPG---VTKKLKRPRPKKKKKYLG---CPAVGEEVD	248
QY	290	QLHOKNEDECAVCRDGGELICDCCPRAFHLACSLPRLREIPSGTWKSSCIQATVDEQ	349
DB	249	GVTLDHODYCEVCOGGEIILCDTCPRAYHLVCLDPELDRAPEKWKSCPHCEKGEVQEA	308
QY	350	PRAEERPRQPPVETPLPPGLRSAGEEVKPPGFLGMDTLVYKKLPAAPSAPLPLGL	409
DB	309	KEEEEEVEE-----GEK-EGEKEE-----DDHMEY-----	334
QY	410	DSALHPLLCVGPQGQNLAPGARCGVCGDGTDLRCTHCAAAFHWCHFPAGTSRPCTG	469
DB	335	-----CRCKDGGELICDCCACISSTHICLNPPLDIPNGE	370
QY	470	LRCRCSGSDVTPAPVEGVL-----APPARLAPGAPKDDTASHEPALRDLESL	519
DB	371	WLCPRCTCVPLKGRVOKILHMWGEPPVAVPAQOAGNP---DVPPPRPLQGRSSEFF	427
QY	520	LSHPTFDGIQW	531
DB	428	VR---WVGUSYW	436

RESULT 2

T42761

Bassoon protein - rat

N:Alternate names: brain-specific synapse-associated protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T42761

R:Dieck, S.; Sammarti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex, J. Cell Biol. 142, 499-509, 1998

A:Title: Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively localized

A:Reference number: 22249; MUID:98345363; PMID:9679147

A:Accession: T42761

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3938 <DIE>

A:Cross-references: EMBL:Y16563; NID:93413503; PID:CAN76287.1; PID:93413504

A:Experimental source: strain Sprague Dawley; brain

A:Function:

A:Description: may be involved in cytomatrix organization at the site of neurotransmitter

A:Note: component of the presynaptic cytoskeleton

C:Keywords: coiled coil; zinc finger

Query Match 7.6%; Score 221.5; DB 2; Length 3938;

Best Local Similarity 23.4%; Pred. No. 6.1e-05;

Matches 123; Conservative 57; Mismatches 232; Indels 113; Gaps 25;

QY 104 VDSQPRKGRPRAPVAPKALVPPRLPTKRAS--EENARAAPAL-----TPRGAS 153
 DB 225 MDMTTAPRSKSOQQLSPALSPAHSPKQPLGKEPERSRSPGATQSGPRQAEARATSV 284
 QY 154 PG-----SOLKAPKPKPSSAEQORLPLGNGIQTMSASVQRAVAMSSGGVPGAR 203
 DB 285 PGTQATAPPEVGVSPQPLSTKPTAEPPPPAG-----EAQGSATTPPSGL 333
 QY 204 GAVE---GIIIOVFESGGS---KKCIQVGEFTY-----PSK-----FEDSG--SG 242
 DB 334 GAEEQTQGGTLGKLFGLGASLTQASTLMSVQPEADTQGGSPSKGPRKIVFSDASKAG 393
 QY 243 KNAKSSSGKPLVRKAGAG--AARGGKRLCGQGSVPAPLALPSD-----PQ 290
 DB 394 PRPSSGPGTPTGATKTEPGTPGSGPGLAKTGTGTPSKHGRADHQAASKAAPK 453
 QY 291 LHKNEDECAVC-----RDGGELICDCCP--RAFHACLSPRLREIPGTRCSCC- 340
 DB 454 TMRKERAACPLQAGELNVSGRGRANNTCTACKLRVCTLCGFNTPHLIVETEMLCMCQ 513
 QY 341 ----LQATVQEVQRAEERPPQEVETPLRPLGRLSAGE---EYKGP--PGEPLAGMDTT 391
 DB 514 TKRLLEGLSGEAPR-LPLPTPQEPAGVP---QRAAGASPLKQKQGGPQGGP-----SGS 564
 QY 392 LVYKHLRPPRSAPRLGLDSSALHRLILCYRREGQONLAPARCGVCGDGYLRTCHCAA 451
 DB 565 LPPKASPOAKAKSPQAKKASPOAKKRLRASEPSTSSAPAEKKTGIPVAKAEVPRKPPETA 624
 QY 452 AFHWRCRHPAGTSRPTGL-RCRSCSGDYTPARVEGLVAPSPARLAPGPAKDDTASHERA 510
 DB 625 V-----PPGTTPKAKSGVKRTDPATPVYKVP--EAPKSGEAEPPVKKPSQDLSRPSQ 676
 QY 511 L-----HDDLESLSEHTFEDGILQMAIQSM-----ARPAAPPS 545
 DB 677 LSDTGYSSDGVSSQSEIT--GVVQOEVEQLDSAGVTGPRPPSPS 719

RESULT 3

T34239

hypothetical protein F26F12.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34239

R:Wilson, R.; Bentley, D.; Gatlung, S.

submitted to the EMBL data library, April 1996

A:Description: The sequence of C. elegans cosmid F26F12.

A:Reference number: 221493

A:Accession: T34239

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1829 <MIL>

A:Cross-references: EMBL:U55373; PID:AGC25894.1; GSPDB:GNO0023; CESP:F26F12.7

A:Experimental source: strain Bristol N2; clone F26F12

C:Genetics:

A:Gene: CESP:F26F12.7

A:Map position: 5

A:Introns: 110/3; 441/3; 801/2; 1244/3; 1693/2; 1784/1

Query Match 7.6%; Score 219.5; DB 2; Length 1829;

Best Local Similarity 18.9%; Pred. No. 3.6e-05;

Matches 95; Conservative 56; Mismatches 154; Indels 197; Gaps 15;

QY 22 VDSAFPLHLADHDVYEDKFOETLHKKEGCPQAFHALLSWLTD-----DSTAILD 76
 DB 85 VCAIIGLTVDEFDYD---EEFFQISMLK-----TFSSYIKQLLEANPGTNVSKMYP 134
 QY 77 FWRVLFKDYMLERYGRLPLDSEFPKDVLSQPKKGRPRAPVAPKALVPPRLPTKRASE 136
 DB 135 MEQVYKREYQDHMAAQKPV-----OKQAGSKTPAVSPVPIPPRSAPTKTRGAR 184
 QY 137 EARAAPALTPRGATSPGSQLAKAPKPPRESSAEQORLPLGNGIQTMSASVQRAVAMSS 196
 DB 185 RKRDRSDA-----PDSQD----- 198
 QY 197 GVPVARGAVEGLIQVFEESGSKCIQVGEFTYPSKEDS--GSGKNKARSSGKPL 255
 DB 199 -----FEAFIKQD-----EQLEDDLVKQEDAKIRKARE 229
 QY 256 VRAKGAOGAPRGGGEARLGGQGSVPAPLALPSDQLHKKNEDECAVCRDGGELICDCCP 315
 DB 230 EKKKGLAARAKKAKL--EKGE-----EAENNDVCECKQOGEILLCTCSP 275
 QY 316 RAFHACLSPRLREIPSGTWRSCSLQATVQEVQRAEERPPQEVETPLRPLGRLSAGE 375
 DB 276 RAVHVCIDENMEDEPREGWSCHACIEHGREYK---EEPAKONDE----- 319
 QY 376 EVRGPRGELAGMDTTLVYKHLRPPRSAPRLGLDSSALHRLILCYRREGQONLAPARGC 435
 DB 320 ----- 321
 QY 436 VCGDGTDLRTCHCAAFHWRCHFPAGTSRPTGL-RCRSCSGDYTPARVEGLA----- 489
 DB 322 ICKETENLLDCSCSFPAATCIDPLLEVYKRETWSCPCETVKPEHKIEKILCWKKE 381
 QY 490 -PSPARLAPG--PAKDDTASHER 509
 DB 382 IPYEPLEAGKVASDDAMLKP 403

RESULT 4

B49515

phosphoprotein 75 - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Sep-1997

C:Accession: B49515

R:Kaderelt, S.; Gewert, D.R.; Galabru, J.; Hovanessian, A.G.; Meurs, E.F.

J. Biol. Chem. 268, 24432-24441, 1993

A:Title: Molecular cloning of two new interferon-induced, highly related nuclear phos

A:Reference number: A49515; MUID:94043285; PMID:7693701

A:Accession: B49515

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-371 <RAD>

A:Cross-references: GB:L22343; NID:g402147; PID:g402148

C:Keywords: phosphoprotein

Query Match 7.5%; Score 219; DB 2; Length 371;

Best Local Similarity 25.1%; Pred. No. 7.6e-06;

Matches 68; Conservative 37; Mismatches 110; Indels 56; Gaps 9;

```

QY 89 RYGRLOPLDSEPKVDLSQPKKPPAVPKALVPPRP-----TKRKA 134
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 RKSSEELIDG---TSENNEGRSQKTPSTPRVVGAAASPGHGIGKLOVVKVTKRKD 160
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 -----SEELAAAPALTPRGTAASPGSOLKAKPKKPSSAEQORLPLGNGIQTFMSASVQR 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 DSTNMFVMAVYQKARTKCAKRS-----RSKEKKKKEDICSSSKRRQKNIHRGKTRKD 215
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 AVAMSSGDVPGARGAVEGILLQVFEESGSKKCIQ-VGGEFFYTPSKFEDSGSKNKARSS 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 TVDFHSCSLPTVTCGAKGILYKKMKHGSSVKCIRNEDGTWLPNPEFEVSGKGRN----A 271
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 SGPPPLVAKAGAGAAAGGGEKRLGQGGSVAPALPSPQLHOKNEDECAVCDGGEGLI 309
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 KMKRNKICEG-----MTLGE-----LTKRKNSDECEVCCGGQLT 307
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 CCDCGPRAFLACLSPPLREITPSGTCWRCSSC 340
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 CCGTCPRVFHEDECHIRPV-EAKRMIMSCITFC 337
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5

```

S55259
Tlfl protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jun-2000
C:Accession: S55259
R:Le Douarin, B.; Zechel, C.; Garnier, J.M.; Lutz, Y.; Tora, L.; Pierat, B.; Heery, D.;
EMBO J. 14, 2020-2033, 1995
A:Title: The N-terminal part of Tlfl, a putative mediator of the ligand-dependent active
A:Reference number: S55259; M0ID:95262642; PMID:7744009
A:Accession: S55259
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1051 <LEID>
A:Cross-references: GB:578221; NID:9998814; PIDN:AAB34290.1; PID:9998815
C:Superfamily: Unassigned bromodomain proteins; Bromodomain homology
F:48-136/Domain: RING finger homology <RRN>
F:925-981/Domain: bromodomain homology <BR01>

```

Query Match 7.5%; Score 217; DB 2; Length 1051;

Best Local Similarity 27.2%; Pred. No. 2.9e-05; Mismatches 122; Indels 60; Gaps 15;

```

Matches 85; Conservative 46; Mismatches 122; Indels 60; Gaps 15;
QY 71 STAILDFWVLFKDYNLERYGRLQPLDSEPKVDLSQ---PRKGRKPPAVPKALVPP 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 SSPMIDLAPVGGSYN-----LPSLP-DIDCSSTIMLDIARKDITGDHA---QP 651
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 RLPTKRKASEEARAAPALTPR-GTASPGSOLKAKPP-KKPSSAEQORLPLGNGIQTM 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 RPPSNRTVQ-----SPNSSVSPGLAGPVTMTSVHPIRSPSASVSGRSSG----- 699
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 SASVGRVAMSSGDVPGARGAVEGILLQVFEESGSKK-----CIQVGEFFYTPSKFEDS 239
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 700 -SSSPAGADSTHKVNV--MEPIRIKQ--ENSGPPEYDFPVIVYKQESDESRPONT 754
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 GSGK-----NKARSSSGPKPLVR--AKGAAGAAPG-----GGEARLGQOGSVAP 282
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 755 NYPRRIILTSLLNSSSSSEETVLASDADSTGDQPGILQDENSNGSG--WSDASQKSP 813
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 LALPSDPQLHOKNEDECAVCDGGEGLICDGGCPRAFLACLSPPLREITPSGTCWRCSSC 342
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 814 VHVGETRKEDDNEDECAVCDGGEGLICCEKCPKVFHLTCVPTLTNPSPSEMICITPCRD 873
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 ATVOEVQRAEPP 355
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 874 LKPEVDYCDVP 886
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6

T20160
hypothetical protein T14G8.1 - Caenorhabditis elegans

```

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20160; T24924
R:McMurray, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: T249231
A:Accession: T20160
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <W11>
A:Cross-references: EMBL:267881; PIDN:CAA91798.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone C52G5
R:Matthews, P.
submitted to the EMBL Data Library, November 1995
A:Reference number: T24924
A:Accession: T24924
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <W12>
A:Cross-references: EMBL:267884; PIDN:CAA91810.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone T14G8
C:Genetics:
A:Gene: CESP:T14G8.1
A:Map position: X
A:Introns: 112/3; 453/3; 597/3; 815/2; 1258/3; 1682/2; 1709/3; 1764/1

```

Query Match 7.1%; Score 206; DB 2; Length 1787;

Best Local Similarity 20.4%; Pred. No. 0.00022; Mismatches 105; Conservative 58; Mismatches 147; Indels 204; Gaps 22;

```

QY 17 EIAVANDSAPFLAHALADHIV-VEDEKQFETLHKKEKQCPQAFHALLSWLTLTQ-----D 70
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 EISAAL-----GLTDVDVDEQEEFQSTNLKN-----ESSLVKPYLLKVNPGIN 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 STAILDFWVLFKDYNLERYGRLQPLDSEPKVDLSQPRKGRKPPAVPKALVPPRLPT 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 VTMKPLFLQYKKEF-----QDHMTAHKSI-----QKQKRAKFPV---VPVVPVTPQ 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 KRKASEEARAAPALTPRGTAASPGSOLKAKPKKPSSAEQORLPLGNGIQTMSASVOR 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 EKTI-----PKTRSSARRK----- 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 AVAMSSGDVPGARGAVEGILLQVFEESGSKKCIQVGEFFYTPSKFEDSG-SGKNKARSS 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 -----RDGSG-----EGGHSIDQFEALIKQHEQOQDAEKGEKARIN 231
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 SGPPPLVAKAGAGAAAGGGEKRLGQGGSVAPALPSPDPLHOKNEDECAVCDGGEGLI 309
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 RAAAKYDKRRAALESASAKRAR-KEGQV-----EENHQRN---CEVCNQGDEL 278
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 CCDCGPRAFLACLSPPLREITPSGTCWRCSSCIGATVQEVQRAEPPRQEPVETPLPG 369
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 LCDCTBRVAVACIDENMEQPPREGDWSG-----PACEEHGPVLIVEE----- 321
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 370 LRSAGEEVRPGRPL-AGMPTLVYKHLPLAPPSAALPLGLDSSALHPLLCVGPGEQNL 428
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 322 -----EPAAAND----- 329
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 429 APGARGCVGGGTVALNCTHCAAFHWHCHPAGTSRPTGLRCSSGSDVTPAP---VE 485
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 -----YCRICKETSIIILCDTCTPSSYHAYCIDPPLTEIEGEMWSCRC---IIPERQRIE 382
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 GVLA-----PSPARLAPG--PAKDDTASHEP 509
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 KILSMRWKEISYRPLECKEAGEASKDDVFLKPP 416
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455

R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
 A:Reference number: 214954; MUID:97388474; PMID:9247308
 A:Accession: T03455
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-4957 <PRA>
 A:Cross-references: EMBL:AF010404; NID:92358286; PIDN:AAC51735.1; PID:92358287
 C:Genetics:
 A:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Query Match 6.7%; Score 194.5; DB 2; Length 4957;
 Best Local Similarity 19.9%; Pred. No. 0.0029;
 Matches 117; Conservative 50; Mismatches 173; Indels 247; Gaps 27;

QY 95 PILDSF-----PKD-VDL--SOPRKGRPPAVPKALVPPRLPTKRKASEARAAAP 143
 DB PALDDFSGLDGDTAPLDIDAGSQPEPQPTGSLASELKGSPVL---LDPEELAPVTP 585
 QY 144 AALRP--RGTAAPGSQLAKAPPKPESSAEQRLPLGNGIQTMSAS--VQRAVAMSSGDV 199
 DB MEVYPECKQTAGRGSS-----PCEQEEPRAPVAPTPPTLLIKSDIYNEISNLSGDA 636
 QY 200 PGAGAVEGILIQVFEESGSKKCIQV--GEFTYPSKFE-----DS 239
 DB 637 SASPFGSEPLIGSPDPPEGGSLS-MELGVSTDVSPARDEGSLRLCTDSLPTDDSLCDA 695
 QY 240 GSG-----KNKARSSGPKPLVRAKGA-----OGAPRG--GEARLQGGSV 279
 DB 696 GTATSGKAEGERRRSSPARSRITKQGRSSSPFGRRRPRGAGHGRGRARLAKSTASS 755
 QY 280 PAPLA---LPSDQLHKNED----- 297
 DB 756 IETLVVADIDSSPSKEEEDDDTDMONTVLFSTNDFVLMQMCVCGSFGRAEGHLL 815
 QY 298 -----ECAYCR-----DGELICCGCPRAPFLA 321
 DB 816 ACSQSGCYHPYCVNSKITVMLLKRCVCEIYCEVCGQASDPSRLLCDCCDISYHTY 875
 QY 322 CLSPPLREIPSGTWKRS---SCLOATVQEVOPRAEPPRPOEPVETPLPGLRSAGEVR 378
 DB 876 CLDPPLLTVPKGGKCKWCSCMCGAAS-----PGFHCKMN-- 913
 QY 379 GPRGEPFLAGMDTLVYKHLRPAAPSAPRLPGLDSSALHPLLCVPGEGQONTAPGARCGVG 438
 DB 914 -----SYTHGCPCASLYTGP-----ICHAPVVEEDL----- 939
 QY 439 DGTDLVLRCTHCAAFHWRG-----HPAGTSRPGTGLRCSGSDVTPAPVEGYL 488
 DB 940 -----LIQCRHCEKRMHAGCSLFTEDVDH-----APDEGFQVCSQ---PYVAKPVA 985
 QY 489 ASPPARLAGPAKADTASHEPALHRDLESLSLSEHTFDDGLIOWAIQS 535
 DB 986 PVAPPPELVPMKVEP---EPQYFR-----FEGV--WLTER 1015

RESULT 8
 T03454
 ALR protein - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
 C:Accession: T03454
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
 A:Reference number: 214954; MUID:97388474; PMID:9247308
 A:Accession: T03454
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

A:Residues: 1-5262 <PRA>
 A:Cross-references: EMBL:AF010403; NID:92358284; PIDN:AAC51734.1; PID:92358285
 C:Genetics:
 A:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Query Match 6.7%; Score 194.5; DB 2; Length 5262;
 Best Local Similarity 19.9%; Pred. No. 0.0031;
 Matches 117; Conservative 50; Mismatches 173; Indels 247; Gaps 27;

QY 95 PILDSF-----PKD-VDL--SOPRKGRPPAVPKALVPPRLPTKRKASEARAAAP 143
 DB 835 PALDDFSGLDGDTAPLDIDAGSQPEPQPTGSLASELKGSPVL---LDPEELAPVTP 890
 QY 144 AALRP--RGTAAPGSQLAKAPPKPESSAEQRLPLGNGIQTMSAS--VQRAVAMSSGDV 199
 DB 891 MEVYPECKQTAGRGSS-----PCEQEEPRAPVAPTPPTLLIKSDIYNEISNLSGDA 941
 QY 200 PGAGAVEGILIQVFEESGSKKCIQV--GEFTYPSKFE-----DS 239
 DB 942 SASPFGSEPLIGSPDPPEGGSLS-MELGVSTDVSPARDEGSLRLCTDSLPTDDSLCDA 1000
 QY 240 GSG-----KNKARSSGPKPLVRAKGA-----OGAPRG--GEARLQGGSV 279
 DB 1001 GTATSGKAEGERRRSSPARSRITKQGRSSSPFGRRRPRGAGHGRGRARLAKSTASS 1060
 QY 280 PAPLA---LPSDQLHKNED----- 297
 DB 1061 IETLVVADIDSSPSKEEEDDDTDMONTVLFSTNDFVLMQMCVCGSFGRAEGHLL 1120
 QY 298 -----ECAYCR-----DGELICCGCPRAPFLA 321
 DB 1121 ACSQSGCYHPYCVNSKITVMLLKRCVCEIYCEVCGQASDPSRLLCDCCDISYHTY 1180
 QY 322 CLSPPLREIPSGTWKRS---SCLOATVQEVOPRAEPPRPOEPVETPLPGLRSAGEVR 378
 DB 1181 CLDPPLLTVPKGGKCKWCSCMCGAAS-----PGFHCKMN-- 1218
 QY 379 GPRGEPFLAGMDTLVYKHLRPAAPSAPRLPGLDSSALHPLLCVPGEGQONTAPGARCGVG 438
 DB 1219 -----SYTHGCPCASLYTGP-----ICHAPVVEEDL----- 1244
 QY 439 DGTDLVLRCTHCAAFHWRG-----HPAGTSRPGTGLRCSGSDVTPAPVEGYL 488
 DB 1245 -----LIQCRHCEKRMHAGCSLFTEDVDH-----APDEGFQVCSQ---PYVAKPVA 1290
 QY 489 ASPPARLAGPAKADTASHEPALHRDLESLSLSEHTFDDGLIOWAIQS 535
 DB 1291 PVAPPPELVPMKVEP---EPQYFR-----FEGV--WLTER 1320

RESULT 9
 S58151
 hypochelical protein SPAC2F7.07c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: T38555; S58151
 R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Ralajandram, M.A.; Walsh, S.V. submitted to the EMBL Data Library, July 1995
 A:Reference number: 221799
 A:Accession: T38555
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-607 <GE2>
 A:Cross-references: EMBL:Z50142; NID:91052783; PIDN:CAA90494.1; PID:91052790; GSPDB:G
 A:Experimental source: strain 972h-; cosmid c2F7
 C:Genetics:
 A:Gene: SPAC2F7.07c
 A:Map position: 1

Query Match 6.4%; Score 186.5; DB 2; Length 607;

Best local similarity 22.7%; Pred. No. 0.001;
Matches 83; Conservative 44; Mismatches 132; Indels 107; Gaps 18;

QY 151 TASPSSQAKAPKPKRESSAEQOQLPLGNGIQTMSASVORAVAMSSGDVYGARGAVEGIL 210
DB 169 TAKDLSDISSSMKKANSSK-----PLFSGKLTKFNT-----PPTSEVTENNYY 215
QY 211 IQQVFESGSKKCIQVGGFFTPPSKFEDSGSKNNARSSSGPKPLVRANGAQAAGAGGGE 270
DB 216 TRNVTVYSNOK---HLGNE---SENFNDM---EGRAEDISSNELL----- 251
QY 271 ARLGGQGVAPLALPSPQHLQKNEDECACVRDGGELLCCDCCPRAFHLACLSPURE- 329
DB 252 -----PPPEEY-----TRYNDVYSACHGPNFLCCETCPNSFHTCTCDPIEEK 297
QY 330 -IPSGTWRSCSLQATV-----OEVPRAEPR-----POEPPVEPLLP 368
DB 298 NLPPDAWVCNCKHHSILNEDEGELESNVKEEGTMVVMQLCTYISHNFIQHLPH 357
QY 369 GL-----BSAGEVVRGPPPEPLAGMDTTLVYKHLPAF-----PSADPLPGLDSSALHPLLC 419
DB 358 SISSEFFRCVSGVMGEYIE-----TDVLKHLKSSRRSNGEERDPLL-LKSSGTPIIC 409
QY 420 VGPEQQQLLARGARGVCG-DGTDVLRCTHCAAFHMRCHFRAGTSRPGTGLR---CRSC 475
DB 410 F-----RCHKALVQSILACDYCNMYHDPDLNPLATLP-SNLKMKCPNH 456
QY 476 SGDYTP 481
DB 457 SDHYTP 462

RESULT 10

562475
PBD finger protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000
C:Accession: J37778; S62475
R:Bedcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z21745
A:Accession: J37778
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <BA2>
A:Cross-references: EMBL:Z54366; NID:91019812; PIDN:CAA91193.1; PID:91019817; GSPDB:GNOC
A:Experimental source: strain 972h-; cosmid c16c9
C:Genetics:
A:Gene: SPAC16C9.05
A:Map position: 1L

Query Match 6.2%; Score 179.5; DB 2; Length 404;
Best local similarity 25.5%; Pred. No. 0.0017;
Matches 67; Conservative 27; Mismatches 94; Indels 75; Gaps 12;

QY 294 KNEDECAVCRDGGELICDCCPRAFHLACLSPPL-REIPSGTWRSCSLQATVOEVQPR 351
DB 115 RNVDYCSACGGRLGFLICCEGRCPSFHLSCLEPPLPENIPBSWFCVYC---SISNHP- 170
QY 352 AEPRRQEP-----PVETPLPGLRSAGEV-RGPPG---EPLAGMDTTL 392
DB 171 -----PKNPLSIWSQLYDWIDSONPSQYRLPDDLVHYFHGISRGDTGAVKETEGEDMT- 223
QY 393 VYKHLPAFPAAPRLPGDSSALHPLLCVPRGQQLNARGARGVCGDGD---VLKCH 448
DB 224 --DEFSALPTGSSITNL-----ATCGYCSKPSMGACWVYGCOL 259
QY 449 CAAAFHMRCHFRAGTSRPGTGLRCRSCSGDYTPAPVEGVLAAPPAFLAPGAKDTASH- 507
DB 260 CDFEYHKCKEHA-----KKCSHD--SIGKKGMKRVKKNVIVRTPLVLDITSTNT 306
QY 508 -EPALHDDLESLSEHTFGCIL 529

DB 307 LNPVKMISGMQFLMGEPFSDLL 329

RESULT 11

T42730
Bassoon protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42730
R:Dieck, S.; Sammarti-Vila, L.; Langreese, K.; Richter, K.; Kindler, S.; Soyke, A.; W.
J. Cell Biol. 142, 499-509, 1998
A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively local1
A:Reference number: Z22249; MUID:98345363; PMID:9679147
A:Accession: T42730
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3942 <DIE>
A:Cross-references: EMBL:Y17034; NID:93413809; PIDN:CAA76598.1; PID:93413810
A:Experimental source: strain 129 SVJ
C:Genetics:
A:Map position: 9P1
A:Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
A:Note: bassoon
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmi
A:Note: component of the presynaptic cytoskeleton
C:Keywords: coiled coil; zinc finger

Query Match 6.2%; Score 178.5; DB 2; Length 3942;
Best local similarity 22.1%; Pred. No. 0.02;
Matches 118; Conservative 50; Mismatches 230; Indels 135; Gaps 24;

QY 101 PKQVDSLQPRKGRKPRPAVKALVPPRLPTRKKAASEKRAAPALTRGTASPG-SQLK 159
DB 250 PAKQPLKPRDQKSPRG-PGATQSGPR-----QEAARATSVPGPT-QATPAPEVGRVS 301
QY 160 AKPPKPPRESSAEQOQLPLGNGIQTMSASVQRAVAMSSGDVYGARGAVEGILIQVFESG 219
DB 302 PQPLSLTKPSTAEERRPAG-----EAQGSATTVSGIAGAGT-----QBSL 344
QY 220 SKKCIQVGEFT-----PSK-----FEDSG--SGKNKARSSGP 252
DB 345 TGLKFLGIGASLITQASTLMSVQPADTQGGSPSPKGPPIYFSDASKRAGRPSPGSGP 404
QY 253 KPLVRAAGAGA--APGGEARLQOQGSVVP-----IALPSDPLQKNEDEC 299
DB 405 GPPPGAKTEPGARTPGSPGALAKTGCTAPSKGRARHQAASRAAKPKTMPKERRAAC 464
QY 300 AVC-----RDGELICDCCP-RAVHLACLSPPLREIPSGTWRSCSS-----LQNTV 345
DB 465 PLCOAEIIVSGSKRGANNTCTACKLQVCNLCGFNFTPLVKTETWLCNCTKRLLESL 524
QY 346 QEVQRAEERPPQEPVETPLPGLRSAGE---EVRGPPG-----EPLAGMDT 390
DB 525 GEPAP-LPLPTPOQPPAGVP---HRAAGAPRLQKQKGGQIGQPSGLPAKASQATQAS 579
QY 391 TLVYKHLPAFPAAPRLPGDSSALHPLLCVPRGQQLNARGARGVCGDGTDLVLCYHCA 450
DB 580 PQATKASPOATKASPOATKASPOAKPLRATERSTSSAQAQKATATPAKAEVVKPP- 636
QY 451 AAFHMRCHFRAGTSRPGTGLRCRSCSGDYTPAPVEGVLAAPSP-----ARLAPRAKD 502
DB 637 -----PETTVPPETP---KAKSVKKTDPAPVAVKPEAPKGEAEPEVKPYSO 684
QY 503 DTASHEPAL-----HRDDESLSEHTFGDILQWALISM-----ARPAAPPS 545
DB 685 DLSRSPQSLSDTGYSSDGVSSQSEIR--GVYQGEVVDQDSAGTVGTRPSPS 735

RESULT 12

102518
hypothetical protein At2g37520 [Imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein F13M22.2

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C:Accession: T02518; G84793
 R:Rounsley, S.D.; Kaul, S.; Liu, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, June 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.
 A:Reference number: 214677
 A:Accession: T02518
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-825 <R0U>
 A:Cross-references: EMBL:AC004684; NID:g3236234; PID:g3236235
 A:Experimental source: Cultivar Columbia
 R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Sheen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.; Nause, D.; Nienman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84793
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-825 <SNO>
 A:Cross-references: GB:AE002093; NID:g3236235; PID:AMC23623.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: F13M22.2; At2g37520
 A:Map position: 2
 A:Intons: 189/3; 212/3; 282/3; 304/1; 366/2; 388/2; 415/3; 433/2; 489/1; 550/2; 569/3;
 C:Superfamily: Arabidopsis thaliana hypothetical protein F4P12.380

Query Match
 Best Local Similarity 34.1%; Score 171; DB 2; Length 825;
 Matches 29; Conservative 18; Mismatches 36; Indels 2; Gaps 1;

QY 295 NEDECAVCRGGEICDCCPRRAHACLSPLEIRSGTSCSCLOATVQVQRAAE 354
 Db 463 SDKMCSTCGGGDLICAGCPQAHATCLK--FQSMPEGTWYSCSCNDGPISKKATTTD 520
 QY 355 PRQPEPPVETPLPGLRSAGEEVG 379
 Db 521 PSCNARPIVRLSRVYKAPESDIOG 545

RESULT 13
 QOBE3
 BHLF1 protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
 C:Accession: A03742
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
 M01. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6092825
 A:Accession: A03742
 A:Molecule type: DNA
 A:Residues: 1-660 <BAN>
 R:Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation: protein coding region
 C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52 C:Superfamily: human herpesvirus 4 BHLF1 protein

Query Match
 Best Local Similarity 23.4%; Score 168.5; DB 1; Length 660;
 Matches 120; Conservative 26; Mismatches 192; Indels 175; Gaps 25;

QY 109 PRKGRP-----PAVPALVPPLPTKRKASEARAAAPALPR-----GTASPGS 156
 Db 172 PGAGRGPGGPTGGRPAEPA-----PCTPA--APGCGGAAPVPGATPHRPGSGPADPPA 225
 QY 157 QLKAKP-----PKKPSSAEQORLPLGNGIQTMSASVQRAVAMSSGVPGAR----- 203

Db 226 AARLPPEQERLPODLAAACRCPAPPTTSGAAAOQTHRRPCCPSARNPCCPTWR 285
 QY 204 ---GAVEGIILOQVFEESGSKKCIQVGGFEFTPSKFEEDSGGKNKARSSGKPLVRAG 260
 Db 286 RRSQAQKG-----HP-----PQAGQ--RPSPTGGR--AAPG 315
 QY 261 AQC--AAPGGEARLGOOGSVAPL--ALPSDPLHOKNEDECAVCRDGGELICDGGPR 316
 Db 316 APGTPAAPCGGGAAPVSGATPHRERGSGRPADPAAALPREQERLPODLAAAOQCPA 375
 QY 317 AFHLACSLPLR-----EIPSG-----TWCSCLQATVQEVQRAEP 355
 Db 376 G-----PPTRSCGAANOQTHRRPCCPSARNPCCPTWRHRSQAQ----- 416
 QY 356 RPOEPVETPLPGLRSAGEEVGRPREPLAGMDTTLVYKILPAPPSAARPLGDSALH 415
 Db 417 RGNRPQAGGRPSGPTGGRPAARCAETPA-----PPGGAAPV--SGATPH 463
 QY 416 PLLCVG-----PEGO-----QNLAPGARC-----GYCGDGTD----- 442
 Db 464 PERGSGPADPRAAARLPREQERLPODLAAAOQCPAPPTTSGAAAOQTHRRPCCPR 523
 QY 443 VLKCTHCAAFHNRCHRPACTSRPSTGLRCRSCG-----DVTAPVEGLAPS 491
 Db 524 SANPCCPRTWRRRSGAQRHNPAGQRPSCPTGGRPAARCAETPAARPGGGAAPVS 583
 QY 492 PARLAP-----GPAKDDTAS-----HEPALHRD 514
 Db 584 GATPHRERGSGRPADPRAAALPREQERLPOD 616

RESULT 14
 T45070
 protein kinase homolog R31240.1 [Imported] - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T45070
 R:Lamerdin, J.; McCreedy, P.; Stillwagen, S.; Ramirez, M.; Carrano, A.
 submitted to the EMBL Data Library, November 1996
 A:Description: Characterization by genomic sequence analysis of a gene-rich
 A:Reference number: 222906
 A:Accession: T45070
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1237 <LAM>
 A:Cross-references: EMBL:AD000092; PID:AA851171.1
 A:Experimental source: cell line 5HL2-B; fibroblast
 C:Genetics:
 A:Map position: 19p13.2
 A:Note: Intron positions not resolved (incomplete sequence)

Query Match
 Best Local Similarity 26.4%; Score 168; DB 2; Length 1237;
 Matches 111; Conservative 41; Mismatches 180; Indels 88; Gaps 22;

QY 116 PAVPALVP-----PRLPTKRKASEARAAAPALPRGTASPSGOLKAKPKPRESS 169
 Db 870 PLSPLAHPPTQASPPPLPGHTVGSSTHTGSPFAKLHSSPPVRRPRASAEPRSPPLK 929
 QY 170 AEOGRPLGNGIQTMSASVQRAVAMSSGDV--PGARGAVEG--ILIQVFEESGSKKCIQV 226
 Db 930 RVQSAEKLG---ASLSADKKGALRKHSLVGHPRDKFNGELALHSLAESDGEPRVVG 986
 QY 227 GGEFTPSKFEEDSGGKNKARSSGKPLVRAKGA-----QGAAPGGGARLGOQS 278
 Db 987 LG---APQVAVNRKRGRESPLSGADPLL--PEGASRPVSKKEESPAGACATPPRAT 1042
 QY 279 VPADLAPSD--OLHOKNEDECAVCRDGGELICDGGPRAPHLACLSPLEIRPSGTWR 336
 Db 1043 TPGRRTLEVDYGCIRHGSVQRE-----DG-----TGMARAVAKAALS--PVQENETG--R 1089
 QY 337 CSSCLQATVQEVQRAEPREPPVETPLPGLRSAGEEVGR--PGEPLA--GMD--- 389

Db 1090 RSSSEAGTPIVIVVEAPREAKAV-VPOPLGADSKGLQEPAPLAPSVPEAPRGRRMV 1148
QY 390 -----TTLV-YKHLPAAPSAAPLPGLDSSALHPLLCVPEGOQNLAPAGRCVCGDGT 441
Db 1149 LEVVEERTTLGSPRKPSPKSPSEP--QTPSLAPAKCASPSSAVTPVPPA--SLUGSGT 1204
QY 442 DVLRCTHCAAFHWRCHPACTSRPGTGLRCRSCGSDVTAPV--EGVLAPSPARLAPGP 499
Db 1205 -----KPOVGLTSRCPAEAVPPAGLTKKGVSSPAP----PGP 1237

RESULT 15

S52835

hypothetical protein YMR075w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR075w

C:Species: Saccharomyces cerevisiae

C>Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C/Accession: S52835

R/Pearson, D.; Bowman, S.
Submitted to the EMBL Data Library, April 1995

A/Reference number: S52814

A/Accession: S52835

A/Molecule type: DNA

A/Residues: 1-684 <PEP>

A/Cross-references: EMBL:248952; NID:9763008; PIDN:CAA88800.1; PID:9763022; GSPDB:GN0001

A/Experimental source: strain AB972

C/Genetics:

A/Gene: MIPS:YMR075w

A/Cross-references: SGD:S0004680

A/Map position: 13R

Query Match

Best Local Similarity 5.5%; Score 161; DB 2; Length 684;

Matches 48; Conservative 34; Mismatches 85; Indels 54; Gaps 8;

QY 294 KNEDECAVCRDGGELICDGCPRAPHLACLSPPL--REIPSGTWRCSSC-----LQA 343
Db 258 ENEDFCACNOSGSLCCDCTCPKSFHCLDPLDPDNNLPKGDWHCNCKFKIFINNSMA 317
QY 344 TVQEVOPRAEPR-----POEPVETPLPPGLRSAGEEVR--GPPGE----- 383
Db 318 TLKKIESNFIKONNNVAFKLLFNIDSHNPKQPLPNYIKETFPAAVKTGSRGQYSDEND 377
QY 384 --FLAG--MDTTLVYKHLPAAPSAAPLPGLDSSALHPLLCVPEGOQNLAPAGRCVCGD 439
Db 378 KIPLDTRQLFNTSYGOSITKLDSYNDPTIHIDNSGKFLICY-----KCNQTRL 425
QY 440 GT-----DVLRCTHCAAFHWRCHPACTSRPGTGLRC 472
Db 426 GSNMHPENSRLIMTCDYQTPWHLDCVPRAFSKMLGSKMKC 466

Search completed: March 13, 2003, 17:54:29
Job time : 30 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:45:58 ; Search time 32 Seconds
(without alignments)
3509.240 Million cell updates/sec

Title: us-09-509-595b-2

Perfect score: 2902 1 MATDAAALRLRLRLHREIAV.....DGLIOWAIGSMARPAAPPS 545

Sequence: 1 MATDAAALRLRLRLHREIAV.....DGLIOWAIGSMARPAAPPS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2687	92.6	515	4 075745	075745 homo sapien
2	2005	69.1	552	11 0920E3	0920E3 mus musculu
3	1991	68.6	548	11 091W9	091W9 mus musculu
4	1989.5	68.6	551	11 091X0	091X0 mus musculu
5	1975.5	68.1	547	11 091W8	091W8 mus musculu
6	1853.5	63.9	493	11 091W7	091W7 mus musculu
7	1839.5	63.4	489	11 091W5	091W5 mus musculu
8	1838	63.3	492	11 091W6	091W6 mus musculu
9	1824	62.9	488	11 091W4	091W4 mus musculu
10	1438.5	49.6	409	11 091W3	091W3 mus musculu
11	1424.5	49.1	405	11 091W1	091W1 mus musculu
12	1423	49.0	408	11 091W2	091W2 mus musculu
13	1409	48.6	404	11 091W0	091W0 mus musculu
14	270.5	9.3	2000	4 09Y410	09Y410 homo sapien
15	253	8.7	1954	4 08TD10	08TD10 homo sapien
16	229	7.9	689	4 09HB58	09HB58 homo sapien

17	221.5	7.6	3938	11 088778	088778 ratu norv
18	219.5	7.6	1829	5 019815	019815 caenorhabdi
19	219	7.5	408	4 014977	014977 homo sapien
20	217.5	7.5	1214	4 09B004	09B004 homo sapien
21	217	7.5	516	11 08R154	08R154 mus musculu
22	214	7.4	429	11 09D108	09D108 mus musculu
23	214	7.4	1209	11 0924W6	0924W6 mus musculu
24	210.5	7.3	245	4 09H930	09H930 homo sapien
25	209	7.2	885	4 096795	096795 homo sapien
26	197.5	6.8	5038	11 090YX7	090YX7 mus musculu
27	196	6.8	4957	4 014687	014687 homo sapien
28	194.5	6.7	5262	4 014686	014686 homo sapien
29	194.5	6.7	680	4 0968B5	0968B5 homo sapien
30	192	6.6	5085	11 09JXK5	09JXK5 ratu norv
31	188	6.5	888	5 09M410	09M410 drosophila
32	187.5	6.5	1139	5 09VDS5	09VDS5 drosophila
33	186.5	6.4	704	4 0960T6	0960T6 homo sapien
34	186	6.4	878	5 08T001	08T001 drosophila
35	184.5	6.4	3851	4 043161	043161 homo sapien
36	184.5	6.4	3926	4 09UPA5	09UPA5 homo sapien
37	184.5	6.4	2228	10 046579	046579 arabidopsis
38	184	6.3	356	11 08VEC8	08VEC8 mus musculu
39	182.5	6.3	810	10 0945C8	0945C8 zea mays (m
40	182.5	6.3	5147	4 09Y6V0	09Y6V0 homo sapien
41	182	6.3	627	5 08SYJ8	08SYJ8 drosophila
42	181	6.2	887	5 016102	016102 drosophila
43	181	6.2	1133	5 09U377	09U377 drosophila
44	180	6.2	812	10 09MB44	09MB44 oryza sativ
45	179.5	6.2	10		

ALIGNMENTS

RESULT 1	ID	075745	PRELIMINARY;	PRT;	515 AA.
AC	075745;				
DT	01-NOV-1998 (TREMBLrel. 08, Created)				
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	AIRe protein.				
GN	AIRe.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-98061087; PubMed-9398840;				
RA	Aaltonen J., Björns P., Perheentupa J., Horelli-Kuitunen N.,				
RA	Palot A., Peltonen L., Lee Y.S., Francis F., Hennig S., Thiel C.,				
RA	Lehrach H., Yaspo M.L.;				
RT	"An autoimmune disease, APECED, caused by mutations in a novel gene				
RT	featuring two PHD-type zinc finger domains.";				
RL	Nat. Genet. 17:399-403(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Lee Y.S., Francis F., Hennig S., Thiel C., Reinhard R., Lehrach H.,				
RA	Yaspo M.L.;				
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ009610; CA08759.1; -				
DR	InterPro; IPR000770; SAND_domain.				
DR	InterPro; IPR004865; SPI00.				
DR	InterPro; PFO0628; PHD; 1.				
DR	Pfam; PF01342; SAND; 1.				
DR	Pfam; PF01372; SPI00; 1.				
DR	SMART; SM00249; PHD; 1.				
DR	SMART; SM00258; SAND; 1.				
SO	SEQUENCE 515 AA; 54600 MW; 2AF5B3DC28605A3A CRC64;				

Query Match

92.6%; Score 2687; DB 4; Length 515;

Best Local Similarity 94.5%; Pred. No. 3.8e-174;
Matches 515; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

```

QY 1 MATDALARLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFH 60
Db 1 MATDALARLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFH 60
QY 61 ALLSWLTLQDSTALIDFWRLFKDYNLERGLPILDSFPKDVDSQPRKGRKPPAVPR 120
Db 61 ALLSWLTLQDSTALIDFWRLFKDYNLERGLPILDSFPKDVDSQPRKGRKPPAVPR 120
QY 121 ALVPPRLPTKRKASEEARAAPALTPRGTAAPSGQLKAKPKKRESSAEQRLPLGNG 180
Db 121 ALVPPRLPTKRKASEEARAAPALTPRGTAAPSGQLKAKPKKRESSAEQRLPLGNG 180
QY 181 IOTMSASVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSG 240
Db 181 IOTMSASVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSG 240
QY 241 SGKRRSSSGPRPLVRKAGAGAPGGEARLGGQGVPAFLALPSDQLHKNDECA 300
Db 241 SGKRRSSSGPRPLVRKAGAGAPGGEARLGGQGVPAFLALPSDQLHKNDECA 300
QY 301 VCRBGELICDGCPRAFHCLSPPLREIPSGTWRCSSCIQATVQEVOPRAEPPROP 360
Db 301 VCRBGELICDGCPRAFHCLSPPLREIPSGTWRCSSCIQATVQEVOPRAEPPROP 360
QY 361 PVETPLPGRSAGEEVRGPEPLAGDITLVYKHLPAAPSADPLGLDSSALHPLLCV 420
Db 361 PVETPLPGRSAGEEVRGPEPLAGDITLVYKHLPAAPSADPLGLDSSALHPLLCV 420
QY 421 GREGQONLARGANGCV-----W-----TGLCRSCSGVY 480
Db 421 GREGQONLARGANGCV-----W-----TGLCRSCSGVY 480
QY 481 PAVEGVLAESPRLAPGPAKDDTASHEPALHRDLESLSHPTFDGILLQMAISMAPPA 540
Db 481 PAVEGVLAESPRLAPGPAKDDTASHEPALHRDLESLSHPTFDGILLQMAISMAPPA 540
QY 541 APPPS 545
Db 541 APPPS 515

```

RESULT 2

```

QY 0920E3 PRELIMINARY: PRT: 552 AA.
AC 0920E3:
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Aire protein (AUTOIMMUNE regulator).
GN Aire.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Karin B., Schweiger M., Wertz K., Poulson R., Christensen H.M.,
RA Rosenthal A., Lehtach H., Yaspo M.L.;
RT "The mouse Aire gene: comparative genomic sequencing, gene
RT organization and expression.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RA MEDLINE=99160890; PubMed=10049735;
RA Miltaz L., Rossier C., Heino M., Petersen P., Krohn K.J.E., Gos A.,
RA Morris M.A., Shimizu N., Antonarakis S.E., Scott H.S.;
RT "Isolation and characterization of the mouse Aire gene.";
RL Biochem. Biophys. Res. Commun. 255:483-490(1999).

```

```

RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=B6; TISSUE=THYMUS;
RX MEDLINE=99168902; PubMed=10049587;
RA Wang C.Y., Shi J.D., Davoodi-Sermitoni A., She J.X.;
RT "Cloning of Aire, the mouse homologue of the autoimmune regulator
RT (AIRE) gene responsible for autoimmune polyglandular syndrome type 1
RT (ASPI).";
RL Genomics 55:322-326(1999).
RN 14
RP SEQUENCE FROM N.A.
RA Halonen M., Peltto-Huikko M., Palvimo J., Björnsen P., Peltonen L.,
RA Urtanen I., Kolmer M.;
RT "Expression of the mouse Aire.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN 15
RP SEQUENCE FROM N.A.
RA Shi J.D., Wang C.Y., Marron M.P., Ruan Q.Q., Huang Y.Q., Dettler J.C.,
RA She J.X.;
RT "Chromosomal Localization and Complete Genomic Sequence of the Murine
RT Autoimmune Regulator Gene (Aire).";
RL Autoimmunity 0:0-0(1999).
RN 16
RP SEQUENCE FROM N.A.
RC STRAIN=MOD, AND SJL;
RA Shi J.D., Wang C.Y., Marron M.P., Ruan Q.Q., Huang Y.Q., Dettler J.C.,
RA Davoodi-Sermitoni A., She J.X.;
RT "Complete genomic sequence, gene structure and localization of the
RT mouse Aire gene.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ132243; CAB36909.1; -
DR EMBL: AJ007715; CAA07620.1; -
DR EMBL: AF079536; AAD20444.1; -
DR EMBL: AJ243821; CAB66141.1; -
DR EMBL: AF105002; AAD46421.1; -
DR EMBL: AF128773; AAF36482.1; -
DR EMBL: AF128772; AAF36481.1; -
DR MGI: MGI:1338803; Aire.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001965; ZnF-PHD.
DR InterPro: IPR001841; ZnF-ring.
DR Pfam: PF01342; SAND; 1.
DR Pfam: PF03172; Sp100; 1.
DR SMART: SM00001; EGF-like; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR SMART: SM00258; SAND; 1.
SQ SEQUENCE 552 AA; 59042 MW; BF30F66B71239A CRC64;

```

Query Match 69.1%; Score 2005; DB 11; Length 552;

Best Local Similarity 71.8%; Pred. No. 6.3e-128;

Matches 393; Conservative 37; Mismatches 108; Indels 10; Gaps 6;

```

QY 4 DAALRLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFHALL 63
Db 5 DAALRLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFHALL 64
QY 64 SWLTLQDSTALIDFWRLFKDYNLERGLPILDSFPKDVDSQPRKGRKPPAVPRKALV 123
Db 65 SWLTLQDSTALIDFWRLFKDYNLERGLPILDSFPKDVDSQPRKGRKPPAVPRKALV 124
QY 124 PPRPLPTKRKASEEARAAPALTPRGTAAPSGQLKAKPKKRESSAEQRLPLGNGIQT 183
Db 125 PPRPLPTKRKASEEARAAPALTPRGTAAPSGQLKAKPKKRESSAEQRLPLGNGIQT 184
QY 184 MSASVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSGSG 242
Db 185 MSASVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSGNL 244
QY 243 KNRARSSSGPRPLVRKAGAGAPGGEARLGGQGVPAFLALPSDQLHKNDECAVC 302

```

```

Db 245 KKKARSSGSLKRVYAKAGQVITIPGDEQKVGQCGVPLPSPSEQVQNKNEDECAYC 304
QY 303 RDGELICDCGCPRAFHILACISPLPRLPISGTCWRCSSCLOATVOEVO PRAEPPROEPPV 362
Db 305 HDGGLICDCGCPRAFHILACISPLPRLPISGTCWRCSSCLOATVOEVO PRAEPPROEPPV 364
QY 363 ETPPLPGISAGEEVRGPPGEPLAGMDTLVYKHLAPRPSAARPLPGIDSSALHILCVGP 422
Db 365 ETPPLVGLRSASEKTRGSPRELKASSDAAVTYVNLAPHPAAPL--LEPSALCPLLSAGN 422
QY 423 EGOQNLAPGARCGVCGDGTDLVLRCTHCAAFHWRCHFPAGTSRPGTGLRSCSGDYTPA 482
Db 423 EGRGCPAPARSACVCGDGTDLVLRCAHCAAFHWRCHFPPTAARPTNLRCCKSCSADSTPT 482
QY 483 P-VEGVLP-SPARLAPGPAK--DDTASHEPALHRDLESLSLSEHTFDGILQWAIQSMAR 538
Db 483 PGTGGEAVPTSGPRAPGLAKVGDSDASHDPVLHRDLESLSLSEHTFDGILQWAIQSMAR 542
QY 539 PAA---PPPS 545
Db 543 PLAETPPSS 552

```

RESULT 3

```

O9JLM9 PRELIMINARY; PRT; 548 AA.
AC 09JLM9;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6;
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator
RT gene (Aire)";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFI28116; AAF36461.1; -.
DR MGD; MGI:1338803; Aire.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000770; SAND_domain.
DR InterPro; IPR004865; SPI00.
DR InterPro; IPR001965; ZnF_PHD.
DR InterPro; IPR001841; ZnF_ring.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF01342; SAND; 1.
DR Pfam; PF03172; SPI00; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00184; RING; 2.
DR SMART; SM00258; SAND; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 548 AA; 58631 MW; 0B0524A3CEDCB6 CRC64;

```

Query Match 68.6%; Score 1991; DB 11; Length 548;
 Best local similarity 71.6%; Pred. No. 5, 5e-127;
 Matches 394; Conservative 37; Mismatches 105; Indels 14; Gaps 7;

```

QY 4 DAARLRLRLRTEIAVAVDSAPFLIHALADHDVYEDKFOETLHLKEKGCQAFHALL 63
Db 5 DGMRLRLRLRTEIAVAVDSAPFLIHALADHDVYEDKFOETLHLKEKGCQAFHALL 64
QY 64 SWLITDSTALIDFWVLEKDVNLERYGRLOPILDSFPKDVDSOPRKGKPPAVPKALV 123
Db 65 SWLITRSGAILDWRILFKDYINLERYSRHSISILDSGFPKDVDSOPRKGKPPAVPKALV 124

```

```

QY 124 PPPRLPTKKRASEARAAAPALTPRGTAHSQULAKAPPKPKESSARQRLPLNGIOT 183
Db 125 LPPRPPTKKRALDEPRATPATLASKSVSPGSHLKKPKPKPDONLESOMLPLDNGIOT 184
QY 184 MSASVQRAVAMSSGDVAGAVGEGILIQVFEESGSKCIQVGEFTYPSKFEED-SGSG 242
Db 185 MASVQRAVAVASGDVGTGAVGEGILIQVFEESGSKCIQVGEFTYPSKFEEDSGNL 244
QY 243 KKKARSSGSKPLVYAKAGQVITIPGDEQKVGQCGVPLPSPSEQVQNKNEDECAYC 302
Db 245 KKKARSSGSLKRVYAKAGQVITIPGDEQKVGQCGVPLPSPSEQVQNKNEDECAYC 300
QY 303 RDGELICDCGCPRAFHILACISPLPRLPISGTCWRCSSCLOATVOEVO PRAEPPROEPPV 362
Db 301 HDGGLICDCGCPRAFHILACISPLPRLPISGTCWRCSSCLOATVOEVO PRAEPPROEPPV 360
QY 363 ETPPLPGISAGEEVRGPPGEPLAGMDTLVYKHLAPRPSAARPLPGIDSSALHILCVGP 422
Db 361 ETPPLVGLRSASEKTRGSPRELKASSDAAVTYVNLAPHPAAPL--LEPSALCPLLSAGN 418
QY 423 EGOQNLAPGARCGVCGDGTDLVLRCTHCAAFHWRCHFPAGTSRPGTGLRSCSGDYTPA 482
Db 419 EGRGCPAPARSACVCGDGTDLVLRCAHCAAFHWRCHFPPTAARPTNLRCCKSCSADSTPT 478
QY 483 P-VEGVLP-SPARLAPGPAK--DDTASHEPALHRDLESLSLSEHTFDGILQWAIQSMAR 538
Db 479 PGTGGEAVPTSGPRAPGLAKVGDSDASHDPVLHRDLESLSLSEHTFDGILQWAIQSMAR 538
QY 539 PAA---PPPS 545
Db 539 PLAETPPSS 548

```

RESULT 4

```

O9JLX0 PRELIMINARY; PRT; 551 AA.
AC 09JLX0;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6;
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator
RT gene (Aire)";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFI28115; AAF36460.1; -.
DR MGD; MGI:1338803; Aire.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000770; SAND_domain.
DR InterPro; IPR004865; SPI00.
DR InterPro; IPR001965; ZnF_PHD.
DR InterPro; IPR001841; ZnF_ring.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF01342; SAND; 1.
DR Pfam; PF03172; SPI00; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00184; RING; 2.
DR SMART; SM00258; SAND; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 551 AA; 58913 MW; 77C75E773B48B72C CRC64;

```

Query Match 68.6%; Score 1989.5; DB 11; Length 551;

Best Local Similarity 71.6%; Pred. No. 7e-127;
Matches 394; Conservative 37; Mismatches 108; Indels 11; Gaps 7;

QY 4 DAALRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 63
DB 5 DGMRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 64
QY 64 SWLLTODSTALIDFWRVLFKDYNLERYGRLQPLIDSEPKVDLSQPRKRPAPVAKALV 123
DB 65 SWLLTRDSGALIDFWRVLFKDYNLERYGRLQPLIDSEPKVDLSQPRKRPAPVAKALV 124
QY 124 PPPRLPTKRKASEARAAAPALTPRGTAAPGSQLAKAPKPKPPESSAEQORLPLNGIOT 183
DB 125 LPPRPPTKRKALEEPRAATPPTATLASKSVSSPSHLKTPKPKPDGNEQHLPLNGIOT 184
QY 184 MSASVORAVAMSSGDVPGARAGAVEGILIQOVFESGSKKCIQVGEFTYPSKPED-SGSG 242
DB 185 MAASVORAVAVASGDVPGTRAGAVEGILIQOVFESGSKKCIQVGEFTYPSKPEDSGNL 244
QY 243 KKKARSSSGPKPLVRAKGAOGAAGGEGARLGQGGSVAPALAPSDPOLHOKNEDECAVC 302
DB 245 KKKARSSSLKPLVRAKGAOYTITGRDQKVGQCGVPLPSLSEPOVNO-NEDECAVC 303
QY 303 RDGELICDGCPRAFHLACLSPPLREIPSGTWRCSSCIQATVOEVPRAEPRPOEPPV 362
DB 304 HDGELICDGCPRAFHLACLSPLEIPSGTWRCSSCIQATVOEVPRAEPRPOEPPV 363
QY 363 ETPPLPGLRSAGEVRGPRGPELAGMDTLVYKHLPAAPSAAPLPGDSSALHPLLCVGP 422
DB 364 ETPPLVGLRSAGEVTRGPRGPELAGMDTLVYKHLPAAPSAAPLPGDSSALHPLLCVGP 421
QY 423 EGOONLAPGARGCGDGTDLVLRCTHCAAFHMRCHFPAGTSRPGTGRCSSGSDVTPA 482
DB 422 EGRPGAPARSACVCGDGTDLVLRCAHCAAFHMRCHFPAGTSRPGTGRCSSGSDVTPA 481
QY 483 P-VEGVLP-SPARLAPGPAK--DDTASHEPALHRDLSLSEHTFEDGILQMAIOSMAR 538
DB 482 PGTGPAVPTSGPRAPAGLAKVGDSDASHDPVLRHDDLSLSEHTFEDGILQMAIOSMAR 541
QY 539 PAA---PPPS 545
DB 542 PLAEPTPPSS 551

RESULT 5
Q9JLM8 PRELIMINARY: PRT: 547 AA.
AC Q9JLM8:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator gene (Aire).";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF128117; AA036462.1;
DR MGD: MGI:1338603; Aire.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: Pf00628; PHD; 2.

DR Pfam: PF01342; SAND; 1.
DR Pfam: PF03172; Sp100; 1.
DR SMART: SM00001; EGF-like; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR SMART: SM00258; SAND; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN; 1.
SQ SEQUENCE 547 AA; 58503 MW; 421B0BD9A7ACB8C CMC64;
Query Match 68.1%; Score 1975.5; DR 11; Length 547;
Best Local Similarity 71.5%; Pred. No. 6.2e-126;
Matches 393; Conservative 37; Mismatches 105; Indels 15; Gaps 8;

QY 4 DAALRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 63
DB 5 DGMRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 64
QY 64 SWLLTODSTALIDFWRVLFKDYNLERYGRLQPLIDSEPKVDLSQPRKRPAPVAKALV 123
DB 65 SWLLTRDSGALIDFWRVLFKDYNLERYGRLQPLIDSEPKVDLSQPRKRPAPVAKALV 124
QY 124 PPPRLPTKRKASEARAAAPALTPRGTAAPGSQLAKAPKPKPPESSAEQORLPLNGIOT 183
DB 125 LPPRPPTKRKALEEPRAATPPTATLASKSVSSPSHLKTPKPKPDGNEQHLPLNGIOT 184
QY 184 MSASVORAVAMSSGDVPGARAGAVEGILIQOVFESGSKKCIQVGEFTYPSKPED-SGSG 242
DB 185 MAASVORAVAVASGDVPGTRAGAVEGILIQOVFESGSKKCIQVGEFTYPSKPEDSGNL 244
QY 243 KKKARSSSGPKPLVRAKGAOGAAGGEGARLGQGGSVAPALAPSDPOLHOKNEDECAVC 302
DB 245 KKKARSSSLKPLVRAKGAOGR----DEQKVGQCGVPLPSLSEPOVNO-NEDECAVC 299
QY 303 RDGELICDGCPRAFHLACLSPPLREIPSGTWRCSSCIQATVOEVPRAEPRPOEPPV 362
DB 300 HDGELICDGCPRAFHLACLSPLEIPSGTWRCSSCIQATVOEVPRAEPRPOEPPV 359
QY 363 ETPPLPGLRSAGEVRGPRGPELAGMDTLVYKHLPAAPSAAPLPGDSSALHPLLCVGP 422
DB 360 ETPPLVGLRSAGEVTRGPRGPELAGMDTLVYKHLPAAPSAAPLPGDSSALHPLLCVGP 417
QY 423 EGOONLAPGARGCGDGTDLVLRCTHCAAFHMRCHFPAGTSRPGTGRCSSGSDVTPA 482
DB 422 EGRPGAPARSACVCGDGTDLVLRCAHCAAFHMRCHFPAGTSRPGTGRCSSGSDVTPA 477
QY 483 P-VEGVLP-SPARLAPGPAK--DDTASHEPALHRDLSLSEHTFEDGILQMAIOSMAR 538
DB 478 PGTGPAVPTSGPRAPAGLAKVGDSDASHDPVLRHDDLSLSEHTFEDGILQMAIOSMAR 537
QY 539 PAA---PPPS 545
DB 538 PLAEPTPPSS 547

RESULT 6
Q9JLM7 PRELIMINARY: PRT: 493 AA.
AC Q9JLM7:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator gene (Aire).";


```

RESULT 8
O9JLM6 PRELIMINARY: PRT: 492 AA.
ID O9JLM6
AC O9JLM6
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator
gene (Aire).";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF28119; AAF36464.1; -.
DR MGD: MGI:1338803; Aire.
DR InterPro: IPR000345; Cytc_heme_bind.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND_domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001965; ZnF_PHD.
DR InterPro: IPR01841; ZnF_Ring.
DR Pfam: PF00628; PHD; 2.
DR Pfam: PF03142; SAND; 1.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR SMART: SM00258; SAND; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ SEQUENCE 492 AA; 52868 MW; 9BE6428E275E5781 CRC64;

Query Match
Best Local Similarity 63.3%; Score 1838; DB 11; Length 492;
Matches 366; Conservative 31; Mismatches 83; Indels 70; Gaps 8;

OY 4 DAALRLRLRLHRTETIAVAANDSAFPLHLALADHDVVPEDKFOETLHLKEGCGQAFHALL 63
DB 5 DGMRLRLRLHRTETIAVAANDSAFPLHLALADHDVVPEDKFOETLHLKEGCGQAFHALL 64
OY 64 SWLLTODSTAILDFWRLVLEKDYNLERYGLQPLIDSPKDVLSQPKRGRKPPAVKALY 123
DB 65 SWLLTRDSGAILDFWRLVLEKDYNLERYGLQPLIDSPKDVLSQPKRGRKPPAVKALY 124
OY 124 PPPRLPTTKKASSEARAAPALTPRGTSAPSGOLAKAPPKKPESSAEQORLPLGNGIQT 183
DB 125 LPPRPPTTKKALAEPRATPATLASKSVSPGSHLTKTPPKKPDGNGLESQHLPLGNGIQT 184
OY 184 MSASVQRAVAMSSGDVPGARGAVEGILLIQVFESGSKKCIQVGGEFYTPSKFED -SGSG 242
DB 185 MAASVQRAVAVASGDVPGTRGAVEGILLIQVFESGSKKCIQVGGEFYTPNKFEDEPSGNL 244
OY 243 KKKARSSSGPKPLVYRAKGAOGAAPGGEGEARLGGQGSVPAPLALPSPDLQKHNDECAVC 302
DB 245 KKKARSSSGSLKPYVRAKGAQAVITIPGRDEQKVGQCGVPPPLPSLPSEPOVNG -NEDECAVC 303
OY 303 RDGGELICCDGCPRAFHNLACSLPRLREITSGTWKSSCLOATVOEVOAPRAEPPPOEPV 362
DB 304 HDGGELICCDGCPRAFHNLACSLPRLREITSGTWKSSCLOGRVQONLSQGEVSRPPELPA 363
OY 363 ETPLEPGLRSAGEEVRGPGEPLAGMDTTLVKKHLPAFPAAPLPLGDSALHPLLCVGP 422
DB 364 ETP-----GP 368
OY 423 EGGONLAPGARCGVCGDGLVLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSCSGDVTPA 482

```

```

DB 369 -----APSARCSVCGDTEYLKCAHCAAFHMRCHFPAAAPGTNLRCKSCSADSPPT 422
OY 483 P-VEGVLP--SPARLAPGPAK--DDTASHEPALHRDDESLSETPFDGILOMATQSMAR 538
DB 423 PGTPGEAVPTSGPRAPAGLAKVGDSDASHDVLHRDLESLLNHSFPGIILQATQSMAR 482
OY 539 PAA---PPPS 545
DB 483 PLAETPPFSS 492

RESULT 9
O9JLM4 PRELIMINARY: PRT: 488 AA.
ID O9JLM4
AC O9JLM4
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator
gene (Aire).";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF128121; AAF36466.1; -.
DR MGD: MGI:1338803; Aire.
DR InterPro: IPR000345; Cytc_heme_bind.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND_domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001965; ZnF_PHD.
DR Pfam: PF00628; PHD; 2.
DR Pfam: PF03142; SAND; 1.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR SMART: SM00258; SAND; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ SEQUENCE 488 AA; 52457 MW; 628EDC8A863C7355 CRC64;

Query Match
Best Local Similarity 62.9%; Score 1824; DB 11; Length 488;
Matches 365; Conservative 31; Mismatches 80; Indels 74; Gaps 9;

OY 4 DAALRLRLRLHRTETIAVAANDSAFPLHLALADHDVVPEDKFOETLHLKEGCGQAFHALL 63
DB 5 DGMRLRLRLHRTETIAVAANDSAFPLHLALADHDVVPEDKFOETLHLKEGCGQAFHALL 64
OY 64 SWLLTODSTAILDFWRLVLEKDYNLERYGLQPLIDSPKDVLSQPKRGRKPPAVKALY 123
DB 65 SWLLTRDSGAILDFWRLVLEKDYNLERYGLQPLIDSPKDVLSQPKRGRKPPAVKALY 124
OY 124 PPPRLPTTKKASSEARAAPALTPRGTSAPSGOLAKAPPKKPESSAEQORLPLGNGIQT 183
DB 125 LPPRPPTTKKALAEPRATPATLASKSVSPGSHLTKTPPKKPDGNGLESQHLPLGNGIQT 184
OY 184 MSASVQRAVAMSSGDVPGARGAVEGILLIQVFESGSKKCIQVGGEFYTPSKFED -SGSG 242
DB 185 MAASVQRAVAVASGDVPGTRGAVEGILLIQVFESGSKKCIQVGGEFYTPNKFEDEPSGNL 244
OY 243 KKKARSSSGPKPLVYRAKGAOGAAPGGEGEARLGGQGSVPAPLALPSPDLQKHNDECAVC 302
DB 245 KKKARSSSGSLKPYVRAKGAQAGR---DEQKVGQCGVPPPLPSLPSEPOVNG -NEDECAVC 299

```

```

OY 303 RDGELICDGCGRAPFLACISPPLEIPSGTWRCSSCLOATVOEVPRAEPRROEPV 362
DB 300 HDGGLICDGCGRAPFLACISPPLEIPSGTWRCSSCLOATVOEVPRAEPRROEPV 359
OY 363 ETPLPGRKASGEVVRGPEPLAGMDTLVYKHLPAAPSAAPLPGISALHFLCVCGR 422
DB 360 ETP-----GP 364
OY 423 EGOONLAPGARGCVCDDTDLRCHTCAAFHWRCHFPACTSRPGLRCSCSGDYTPA 482
DB 365 -----APSARCSCGDDGTEVLRCACAAAFHWRCHFPCTAARPGTNLRCSCSADSTPT 418
OY 483 P-VEGVAP-SPARLAPGAK---DGTASHEPALHDDLESLSHTHTDGLIQMAIQSMAR 538
DB 419 PGTREAVPTSGPRAPGLAKVGDSDSHDPVLRDLESLSLNEHSPDGLIQMAIQSMAR 478
OY 539 PAA---PPPS 545
DB 479 PLAEPPSS 488

RESULT 10
OYJMW3 PRELIMINARY: PRT: 409 AA.
AC 09JMW3:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Autoimmune regulator.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6;
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator
RT gene (Aire).";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF128122; AAF36467.1; -.
DR MGD; MGI:1338803; Aire.
DR InterPro: IPR000345; Cytochrome_bnd.
DR InterPro: IPR000770; SAND_domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF01342; SAND; 1.
DR Pfam: PF01372; Sp100; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00258; SAND; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 409 AA; 44571 MW; 4F95F691AE91E109 CRC64;

Query Match 49.6%; Score 1438.5; DB 11; Length 409;
Best Local Similarity 76.6%; Pred. No. 1e-89;
Matches 278; Conservative 23; Mismatches 61; Indels 1; Gaps 1;

OY 4 DAALRLRLRLHRTETAVNDSAPFLHALADHDVPEDEKFOETLHLKEGEGCPQAFHALL 63
DB 5 DGMRLRLRLHRTETAVNDSAPFLHALADHDVPEDEKFOETLHLKEGEGCPQAFHALL 64
OY 64 SWLLTQDSALIDFWRLVLFKDYNLERYGRLOPILDSFPKVDLSOPRKGKRPAPKALY 123
DB 65 SWLLTRDGSALIDFWRLVLFKDYNLERYGRLOPILDSFPKVDLSOPRKGKRPAPKALY 124
OY 124 PPRLPTKRKASSEKRAAPALTPRGTAASPSQSLAKAPPKKESSEADQRLPLGNGIQT 183
DB 125 LPPRPPTKRKALEEPRAPPAATLASKSVSSPSHLTKTPPKKEDGNLESQHLPLGNGIQT 184
OY 184 MSASVORAVAMSSGDPVARGAVEGILLIOVFESGSKKCIQVGEFFYTPSKFED-SSSG 242
DB 184 MSASVORAVAMSSGDPVARGAVEGILLIOVFESGSKKCIQVGEFFYTPSKFED-SSSG 242

```

```

DB 185 MAASVORAVYVAGSDVPTGAVGEGILLIOVFESGSKKCIQVGEFFYTPSKFEDPSGUL 244
OY 243 KKKARSSSGPKPLVVRKAGAGAPGGEARLQGOGSVPAPLALPSDPOLHOKNEDECAVC 302
DB 245 KKKARSSSGPKPLVVRKAGAGVITGRBDQKVGQCGVPRPLSPSEPOVQKNEDECAVC 304
OY 303 RDGELICDGCGRAPFLACISPPLEIPSGTWRCSSCLOATVOEVPRAEPRROEPV 362
DB 305 HDGGLICDGCGRAPFLACISPPLEIPSGTWRCSSCLOATVOEVPRAEPRROEPV 364
OY 363 ETP 365
DB 365 ETP 367

RESULT 11
OYJMW1 PRELIMINARY: PRT: 405 AA.
AC 09JMW1:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Autoimmune regulator.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6;
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator
RT gene (Aire).";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF128124; AAF36469.1; -.
DR MGD; MGI:1338803; Aire.
DR InterPro: IPR000345; Cytochrome_bnd.
DR InterPro: IPR000770; SAND_domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF01342; SAND; 1.
DR Pfam: PF01372; Sp100; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00258; SAND; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 405 AA; 44160 MW; 4D820E9642824F3A CRC64;

Query Match 49.1%; Score 1424.5; DB 11; Length 405;
Best Local Similarity 76.3%; Pred. No. 9e-89;
Matches 277; Conservative 23; Mismatches 58; Indels 5; Gaps 2;

OY 4 DAALRLRLRLHRTETAVNDSAPFLHALADHDVPEDEKFOETLHLKEGEGCPQAFHALL 63
DB 5 DGMRLRLRLHRTETAVNDSAPFLHALADHDVPEDEKFOETLHLKEGEGCPQAFHALL 64
OY 64 SWLLTQDSALIDFWRLVLFKDYNLERYGRLOPILDSFPKVDLSOPRKGKRPAPKALY 123
DB 65 SWLLTRDGSALIDFWRLVLFKDYNLERYGRLOPILDSFPKVDLSOPRKGKRPAPKALY 124
OY 124 PPRLPTKRKASSEKRAAPALTPRGTAASPSQSLAKAPPKKESSEADQRLPLGNGIQT 183
DB 125 LPPRPPTKRKALEEPRAPPAATLASKSVSSPSHLTKTPPKKEDGNLESQHLPLGNGIQT 184
OY 184 MSASVORAVAMSSGDPVARGAVEGILLIOVFESGSKKCIQVGEFFYTPSKFED-SSSG 242
DB 185 MAASVORAVYVAGSDVPTGAVGEGILLIOVFESGSKKCIQVGEFFYTPSKFEDPSGUL 244
OY 243 KKKARSSSGPKPLVVRKAGAGAPGGEARLQGOGSVPAPLALPSDPOLHOKNEDECAVC 302
DB 245 KKKARSSSGPKPLVVRKAGAGVITGRBDQKVGQCGVPRPLSPSEPOVQKNEDECAVC 304

```

QY 303 RDGELICDCCPRAFHFLACLSPLREIPSGTWRCSSCIQATVOEQPRAEPRROPEPV 362
 Db 301 HDGELICDCCPRAFHFLACLSPLREIPSGTWRCSSCIQGRVQNLSDPEVSRPELPA 360
 QY 363 ETP 365
 Db 361 ETP 363

RESULT 12

ID Q9JLM2 PRELIMINARY; PRT; 408 AA.
 AC Q9JLM2;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Autoimmune regulator.
 GN AIRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6;
 RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
 RT "Expression and alternative splicing of the mouse autoimmune regulator
 gene (Aire).";
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF128123; AAF36468.1; -;
 DR MGD; MGI:1338803; Aire.
 DR InterPro: IPR000345; Cytc_heme_bind.
 DR InterPro: IPR000770; SAND_domain.
 DR InterPro: IPR004865; Sp100.
 DR InterPro: IPR001965; ZnF_PHD.
 DR Pfam; PF00628; PHD; 1.
 DR Pfam; PF01342; SAND; 1.
 DR Pfam; PF03172; Sp100; 1.
 DR SMART; SM00249; PHD; 1.
 DR SMART; SM00258; SAND; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 408 AA; 44442 MW; 36857622FFED94C8 CRC64;

Query Match 49.0%; Score 1423; DB 11; Length 408;
 Best Local Similarity 76.3%; Pred. No. 1.1e-87;
 Matches 277; Conservative 23; Mismatches 61; Indels 2; Gaps 2;

QY 4 DAALRLRLHRTETIAVAVDSEFRLHALADHDVPEDEKFOETLHLKEKGCQAFHALL 63
 Db 5 DMLRLRLHRTETIAVAIDSAFRLHALADHDVPEDEKFOETLHLKEKGCQAFHALL 64
 QY 64 SWLTDSDAILDFWVLRKDYNLERYGRLOPLDSEPKVDLSQPRKGRKPPAVKALV 123
 Db 65 SWLTRDSAILDFWVLRKDYNLERYGRLOPLDSEPKVDLSQPRKGRKPPAVKALV 124
 QY 124 PPRRLPTKRKASEEAPAAAPALTPRGTAQSGSLAKAPPKPKRESSAEQORLPNGIOT 183
 Db 125 LPPRPPTKRKALEERATPATLASKSVSSPSGHLTKTPPKKPDGWLSEOHLPNGIOT 184
 QY 184 MSASVORAVAMSSGDVPGARGAVEGILIQVFESESGSKKCIQVGEFFYTPSKFED-SGSG 242
 Db 185 MAASVORAVTVAAGDVPGRGAVEGILIQVFESESGSKKCIQVGEFFYTPSKFED-SGSG 242
 QY 243 KKKARSSSPRIIVAKAKAQAAPGGGEARLCOQGSVPAPLALPSDPQLHQKNEDECAVC 302
 Db 245 KKKARSSSGLKPVVAKAKAQAQGR---DBOKVGQDGVPLPSLPSPQVNO-NEDECAVC 303
 QY 303 RDGELICDCCPRAFHFLACLSPLREIPSGTWRCSSCIQATVOEQPRAEPRROPEPV 362
 Db 304 HDGELICDCCPRAFHFLACLSPLREIPSGTWRCSSCIQGRVQNLSDPEVSRPELPA 363
 QY 363 ETP 365
 Db 361 ETP 363

Db 364 ETP 366

RESULT 13

ID Q9JLM0 PRELIMINARY; PRT; 404 AA.
 AC Q9JLM0;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Autoimmune regulator.
 GN AIRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6;
 RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
 RT "Expression and alternative splicing of the mouse autoimmune regulator
 gene (Aire).";
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF128123; AAF36470.1; -;
 DR MGD; MGI:1338803; Aire.
 DR InterPro: IPR000345; Cytc_heme_bind.
 DR InterPro: IPR000770; SAND_domain.
 DR InterPro: IPR004865; Sp100.
 DR InterPro: IPR001965; ZnF_PHD.
 DR Pfam; PF00628; PHD; 1.
 DR Pfam; PF01342; SAND; 1.
 DR Pfam; PF03172; Sp100; 1.
 DR SMART; SM00249; PHD; 1.
 DR SMART; SM00258; SAND; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 404 AA; 44032 MW; F1D712EECA3A7CB CRC64;

Query Match 48.6%; Score 1409; DB 11; Length 404;
 Best Local Similarity 76.0%; Pred. No. 1e-87;
 Matches 276; Conservative 23; Mismatches 58; Indels 6; Gaps 3;

QY 4 DAALRLRLHRTETIAVAVDSEFRLHALADHDVPEDEKFOETLHLKEKGCQAFHALL 63
 Db 5 DMLRLRLHRTETIAVAIDSAFRLHALADHDVPEDEKFOETLHLKEKGCQAFHALL 64
 QY 64 SWLTDSDAILDFWVLRKDYNLERYGRLOPLDSEPKVDLSQPRKGRKPPAVKALV 123
 Db 65 SWLTRDSAILDFWVLRKDYNLERYGRLOPLDSEPKVDLSQPRKGRKPPAVKALV 124
 QY 124 PPRRLPTKRKASEEAPAAAPALTPRGTAQSGSLAKAPPKPKRESSAEQORLPNGIOT 183
 Db 125 LPPRPPTKRKALEERATPATLASKSVSSPSGHLTKTPPKKPDGWLSEOHLPNGIOT 184
 QY 184 MSASVORAVAMSSGDVPGARGAVEGILIQVFESESGSKKCIQVGEFFYTPSKFED-SGSG 242
 Db 185 MAASVORAVTVAAGDVPGRGAVEGILIQVFESESGSKKCIQVGEFFYTPSKFED-SGSG 242
 QY 243 KKKARSSSPRIIVAKAKAQAAPGGGEARLCOQGSVPAPLALPSDPQLHQKNEDECAVC 302
 Db 245 KKKARSSSGLKPVVAKAKAQAQGR---DBOKVGQDGVPLPSLPSPQVNO-NEDECAVC 299
 QY 303 RDGELICDCCPRAFHFLACLSPLREIPSGTWRCSSCIQATVOEQPRAEPRROPEPV 362
 Db 300 HDGELICDCCPRAFHFLACLSPLREIPSGTWRCSSCIQGRVQNLSDPEVSRPELPA 359
 QY 363 ETP 365
 Db 360 ETP 362
 RESULT 14
 ID Q9Y4I0 PRELIMINARY; PRT; 2000 AA.

AC 09y410:
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Zinc-finger heliase.
 GN HZF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98351552; PubMed-9688266;
 RA Aubry F., Mallet M.G., Gallibert F.;
 RT "Identification of a human 17p-located cDNA encoding a protein of the
 Suf12-like heliase family";
 RL Eur. J. Biochem. 254:558-564 (1998).
 DR EMBL: U91543; AAC39923.1; -
 DR InterPro: IPR000953; Chromo.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAH_box.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR InterPro: IPR001965; Znf_PHD.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF00365; Chromo; 1.
 DR Pfam: PF00271; heliase_C; 1.
 DR Pfam: PF00628; PHD; 2.
 DR Pfam: PF00176; SNF2_N; 1.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR PROSITE: PS00598; CHROMO_1; UNKNOWN_1.
 DR PROSITE: PS00513; CHROMO_2; 2.
 DR PROSITE: PS00690; DEAH_ATP_HELICASE; UNKNOWN_1.
 DR ATP-binding; Helicase.
 DR KW
 DR SEQUENCE 2000 AA; 226576 MW; B085A425AC7971F CRC64;
 SQ

Query Match 9.3%; Score 270.5; DB 4; Length 2000;
 Best Local Similarity 23.2%; Pred. No. 8.5e-10;
 Matches 102; Conservative 46; Mismatches 154; Indels 137; Gaps 17;

DB 116 PAVKALVPPRL-----PTKRKASEEARAAAPALTPRGTA SPGSQLKAKPKKPRESSA 170
 123 PGGPALPPPAADIQPPIRAKTKEGK-----GPHKRSKSPRVDPGRK 282
 171 EQQLPLGNGIOTMSASQRAVAMSSGDVPGARGAVEGILLIOVFESGSKKCIQNGEEF 230
 283 KLR-----GKKMAPLKIKL-----GLL-----GCKRK---KGSY 309
 231 YTPS-----KFEDESGSKNKRARSSG--PKPLVRKAGAGAAPGGGEAR-LGGGSPAP 282
 310 VFGSDDEPERPAESDLDGSGVSHASGRPDGVRKTKLRGRGRKKKKVLG---CPAV 365
 283 LALPSDQLKQKNDDECAVCRDGEGLICDGCPRAFHLACLSPRLREIPSGTARSSCLQ 342
 366 AGEEDVGYETDHDQYCEVCGGGEIILCDTPRAYHLVCLDPELDRAPEGKSKSPHCEK 425
 343 ATQVEQVPAEPRPQEPVETPLRPLGRLSAGEEVRGPRGELAGMDTLLVYKHLAPPS 402
 426 EGQWMEKEEVEEYEE-----GEE--EGEKEEE---DDHMEY----- 458
 403 AAPLPGLDSSALHPLLCVGPREGQNLAPARGCVGCGDGDVLRCTHCAAFHNRCHFPAG 462
 459 -----CRVCKDGGELLCCDCACISSYHHCINPPL 487
 463 TSPRGGLRCRSCSGDVTAPAVEGL-----ASPAPRLAPPAKDDTASHPALH 512
 488 PDIPNGEMLCPRCTCPYLKGRVQKILHWRMGEPVAVAPAPQADGPN---DVPPRPLOG 544
 513 RDLLESLSEHTPTGIIQW 531

DB 545 RSEREFEVK---WVGLSYW 560
 RESULT 15
 Q8TD10 PRELIMINARY; PRT; 1954 AA.
 AC Q8TD10;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Chromodomain heliase DNA binding protein 5.
 GN CHD5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId-9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Thompson P.M., Gotch T., White P.S., Brodeur G.M.;
 RT "CHD5, a New Member of the Chromodomain Gene Family, is Preferentially
 Expressed in the Nervous System";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF425231; AAL98962.1; -
 DR Helicase.
 DR KW
 DR SEQUENCE 1954 AA; 223048 MW; E333062B5B5E71F CRC64;
 SQ

Query Match 8.7%; Score 253; DB 4; Length 1954;
 Best Local Similarity 22.6%; Pred. No. 1.3e-08;
 Matches 121; Conservative 61; Mismatches 188; Indels 166; Gaps 25;

QY 40 EDKFOETLHLK-----EGC---POAFHALLS-WLLNQDSTAIIDFMRVLF--RDY 85
 104 KKKKKKARKKKKDEDDDDDDGCKLKPSSGQMAEGLD-----VDY--LFSEEDY 155
 QY 86 N-LERYGRLQPIIDSEPKVDVLSQPRKGRPAVAKALVPPRLPTK-----RKA 134
 156 HTLTNYKAFSQFL-----RPLAKKNPKIPMSKMTV-LGAKKREFSANNPFGS 204
 QY 135 SEEARAAAPALTPRGTA SPGSQLKAKPKKPRESSA DQRLPLGNGIOTMSASQRAVAM 194
 205 SAAAAAANAAYETVTS--PLAVSPQVP-----QVPVI-----RKAKT 244
 QY 195 SSGDVPAGARGAVEGILLIOVFESGSKKCIQNGEEFTYPSKFEDESGSKNKRARSSGPKP 254
 245 KKGKGGVAKKIKG-----SKDGKK---GKKKTAGLKFRRGGISNKKKKSSSEED 294
 QY 225 LVKANGAGAAPGGEARLGGGSPAPLALPSDQLKQ-----NEDECAVCR 303
 295 EREESDFDSASIHASVR---SECSNALGKSKRRKKRKRIDDDGVEYETHDQYCEVQ 350
 QY 304 DGGELICDGCPRAFHLACLSPRLREIPSGTWKSSCLQATQVEQVPAEPRPQEPVE 363
 331 QGGEIILCDTPRAYHLVCLDPELEKAPBEGKWSCPHCEKEIQ--WEKDDDEEEGGE 409
 QY 364 TPLRPLGRSAGEEVRGPRGELAGMDTLLVYKHLAPPSAARLPGLDSSALHPLLCVPE 423
 410 E-----BE-----DDHMEF----- 418
 QY 424 GQQLNAPARGCVGCGDGDVLRCTHCAAFHNRCHFPAGTSPRGGLRCRSCSGDVTAPAP 483
 419 -----CRVCKDGGELLCCDACRPSSYHHCINPRLPEIRNGEMLCPRCTCPRLK 468
 QY 484 VEGVL-----ASPAPRLA-PGPAKDDTASHPALHDLLESLSEHTPTGIIQW 532
 469 VQRIHWRWTEPPAPFMVGLGP-----DVEPSLPPPKPLEGIPREF--FVKWA 516
 DB

Search completed: March 13, 2003, 17:53:32
 Job time : 36 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:45:22 ; Search time 38 Seconds

(without alignments)
1911.094 Million cell updates/sec

Title: US-09-509-595B-2

Perfect score: 2902
Sequence: 1 MARDALRLRLRLHRTETAV.....DGILOAIQSMARPAAPPSS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

_A_Geneseq_101002:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2902	100.0	545	20 AAY06073	Human autoimmu
2	2902	100.0	545	20 AAY01712	A human autoimmu
3	2005	69.1	552	20 AAY06074	Mouse AIRP Protein
4	1433.5	49.1	348	20 AAY01713	A human autoimmu
5	503.5	17.4	254	20 AAY01714	A human autoimmu
6	241	8.3	1911	17 AAR95534	Dermatomyositis sp
7	241	8.3	1912	22 AAM39288	Human polypeptide
8	238	8.2	1936	22 AAM41074	Human polypeptide
9	236.5	8.1	689	23 ABR07636	Human speckled 110
10			351	13 AAR22280	M1-2 antigen clone

11	227	7.8	884	22 ABR71641	Drosophila melanog
12	227	7.6	1050	22 AAM33902	Human polypeptide
13	220.5	7.6	753	23 ABR07637	Human speckled 140
14	213.5	7.4	753	19 AAM57547	Leukocyte specific
15	213.5	7.4	771	22 ABR11257	Novel human diagno
16	212	7.3	663	21 AAY71210	Human irritable bo
17	210.5	7.3	245	22 AAB94440	Human protein sequ
18	204.5	7.0	885	22 ABR11256	Novel human diagno
19	203	7.0	759	22 AAB94613	Human protein sequ
20	192	6.6	582	22 AAM41852	Human polypeptide
21	192	6.6	582	22 AAM41853	Human polypeptide
22	191	6.6	628	22 AAM40067	Human polypeptide
23	187.5	6.5	898	22 ABB60194	Drosophila melanog
24	186.5	6.4	1139	22 ABB61347	Drosophila melanog
25	186	6.4	504	21 AAB42081	Human ORFX ORF1845
26	184	6.3	335	21 AAB42612	Human ORFX ORF2376
27	181	6.2	883	22 ABR11771	Drosophila melanog
28	179.5	6.2	319	22 AAG74399	Human colon cancer
29	179	6.2	605	22 ABR18143	Novel human diagno
30	177	6.1	1413	23 AAE21729	Human PKIN-24 prot
31	175.5	6.0	581	22 AAM40066	Human polypeptide
32	175.5	6.0	634	22 AAB94512	Human protein sequ
33	169	5.8	334	23 ABB89370	Human polypeptide
34	168	5.8	1134	22 AAB00664	Human protein tyro
35	168	5.8	1570	23 ABB87344	Novel human protei
36	167.5	5.8	728	21 AAG20235	Arabidopsis thalia
37	167.5	5.8	777	21 AAG20236	Arabidopsis thalia
38	167.5	5.8	1080	21 AAG31542	Arabidopsis thalia
39	167.5	5.8	1092	21 AAG31541	Arabidopsis thalia
40	167.5	5.8	1138	21 AAG31540	Arabidopsis thalia
41	166.5	5.7	594	21 AAG20237	Arabidopsis thalia
42	166	5.7	549	22 AAB16301	Human novel secret
43	165	5.7	489	22 ABR14285	Novel human diagno
44	164.5	5.7	393	21 AAB28067	Human secreted pro
45	164.5	5.7	393	21 AAB28068	Human secreted pro

ALIGNMENTS

RESULT 1	AAV06073	standard; Protein; 545 AA.
ID	AAV06073	
XX	XX	
AC	AAV06073:	
XX	XX	
DT	16-AUG-1999	(first entry)
XX	XX	
DE	Human autoimmu	polyglandular disease type 1 (APGDI) protein.
XX	XX	
KW	Autoimmu	polyendocrinopathy candidiasis ectodermal dystrophy:
KW	ABECD:	autoimmu polyglandular disease type 1; APGDI; AIRP;
KW	transcription factor;	autoimmu disease; diagnosis; gene therapy;
KW	human.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
FH	Key	Location/Qualifiers
FT	Region	7..11
FT	Region	/note= "LXXLL motif located in putative helical region, signature for nuclear receptor binding"
FT	Region	113..114
FT	Region	/note= "putative bi-partite nuclear localisation"
FT	Region	131..133
FT	Region	/note= "putative bi-partite nuclear localisation"
FT	Domain	299..340
FT	Domain	/note= "PHD zinc finger domain"
FT	Domain	434..475
FT	Domain	/note= "PHD zinc finger domain"
XX	XX	
PN	XX	W09918197-A2.

PD 15-APR-1999.
 XX
 PF 02-OCT-1998; 98WO-EP06294.
 XX
 PR 12-NOV-1997; 97EP-0119810.
 PR 02-OCT-1997; 97EP-0117154.
 PR 08-OCT-1997; 97EP-0117398.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (NAPU) NAT PUBLIC HEALTH INST.
 XX
 PI Aaltonen J, Björnsen P, Horelli-Kuitunen N, Lehrach H;
 PI Palotie A, Peltonen L, Perheentupa J, Yaspo M;
 XX
 DR WPI: 1999-287735/24.
 DR N-PSDB; AAX58605.
 XX
 PT New polypeptide which co-segregates in mutated form - with
 PT autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 XX
 PS Claim 1; Fig 2A; 77pp; English.
 XX
 CC This is the amino acid sequence of human autoimmune polyglandular
 CC disease type 1 (APGD1 or AIRE), as predicted from a composite
 CC sequence (see AAX58605) of isolated cDNA clones and a PCR extension
 CC product. APGD1 protein is a transcription factor or transcription-
 CC associated factor that may associate with vimentin fibres, perhaps
 CC as part of a docking mechanism regulating nuclear translocation.
 CC Aggregates of the mutated protein may prevent formation of vimentin
 CC intermediate filaments. Mutated APGD1 polypeptides co-segregate
 CC with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 CC (APECED). The invention provides vectors and host cells for
 CC preparation of APGD1 polypeptides. A claimed method for testing
 CC for carrier status of APECED, or for the disease state, involves
 CC testing for a mutation in the APGD1 gene or for a mutated form of
 CC the APGD1 polypeptide.
 XX
 SQ Sequence 545 AA;
 Query Match 100.0%; Score 2902; DB 20; Length 545;
 Best Local Similarity 100.0%; Pred. No. 1.5e-211;
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATDALRLRLRLHRTETIAVAVDSAPFLHLALADHDVPEDEKFOETLHKEEGCQPAFH 60
 DB 1 MATDAALRLRLRLHRTETIAVAVDSAPFLHLALADHDVPEDEKFOETLHKEEGCQPAFH 60
 QY 61 ALLSWLLTQDSTAILDFWRVLFKDYNLERYGRLQPIIDSPKDVLSQPRKGRPPAVPK 120
 DB 61 ALLSWLLTQDSTAILDFWRVLFKDYNLERYGRLQPIIDSPKDVLSQPRKGRPPAVPK 120
 QY 121 ALVPPRLPTKRKASEEARAAPALTPRGTAASPSQKAKPKPKPSSAEOQRLPLNG 180
 DB 121 ALVPPRLPTKRKASEEARAAPALTPRGTAASPSQKAKPKPKPSSAEOQRLPLNG 180
 QY 181 IQTMSAVORAAMSSGDVPGARGAVEGILIQVFESGSKCTQVGGSEPTPKFEDSG 240
 DB 181 IQTMSAVORAAMSSGDVPGARGAVEGILIQVFESGSKCTQVGGSEPTPKFEDSG 240
 QY 241 SKKNKRRSSGKRPVLRKAGAGAPGEGARLQGQGSVPAPLALPSDPOLHQNEDCA 300
 DB 241 SKKNKRRSSGKRPVLRKAGAGAPGEGARLQGQGSVPAPLALPSDPOLHQNEDCA 300
 QY 301 WCRDGGELICCGCCPRAFLACISPLRLRETPSGTWRCSSCLQATVQEVQRAEPRPOEP 360
 DB 301 WCRDGGELICCGCCPRAFLACISPLRLRETPSGTWRCSSCLQATVQEVQRAEPRPOEP 360
 QY 361 FVEETPLPGLRSAGEVRSPGEPPLAGMDTTLVYKHLPPAPSAAPLPGDSSALHPLCY 420
 DB 361 FVEETPLPGLRSAGEVRSPGEPPLAGMDTTLVYKHLPPAPSAAPLPGDSSALHPLCY 420
 QY 421 GPEGQONLAPGARGVCGDGTDLRCTHCAAFHWRCHPEPAGTSRPGTGLRCHSCSGDYT 480
 DB 421 GPEGQONLAPGARGVCGDGTDLRCTHCAAFHWRCHPEPAGTSRPGTGLRCHSCSGDYT 480

DB 421 GPEGQONLAPGARGVCGDGTDLRCTHCAAFHWRCHPEPAGTSRPGTGLRCHSCSGDYT 480
 QY 481 PAPVEGLAPSPARLAPGAKDQTASHEPALRDDELSLEHTFPGCIIQMAIOSMARPA 540
 DB 481 PAPVEGLAPSPARLAPGAKDQTASHEPALRDDELSLEHTFPGCIIQMAIOSMARPA 540
 QY 541 APFES 545
 DB 541 APFES 545
 RESULT 2
 ID AAY01712 standard; Protein: 545 AA.
 AC AAY01712;
 XX
 DT 24-JUN-1999 (first entry)
 DE A human autoimmune regulator-1 (AIRE-1) protein.
 KW Autoimmune regulator-1; AIRE-1; immune maturation; immune response;
 KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
 KW APECED; autoimmune polyglandular syndrome type 1; APS I.
 OS Homo sapiens.
 PN WO9915559-A1.
 PD 01-APR-1999.
 PF 23-SEP-1998; 98WO-FI00749.
 PR 23-SEP-1997; 97FI-0003762.
 XX
 PA (FIIM-) FINNISH IMMUNOTECHNOLOGY LTD.
 PI Antonarakis S, Helino M, Krohn K, Kudoh J, Laitio M;
 PI Peterson P, Scott H, Shimizu N;
 DR WPI: 1999-244390/20.
 DR N-PSDB; AAX26936.
 XX
 PT Autoimmune regulator 1 (AIRE) DNA sequence
 PS Claim 5; Page 24-26; 59pp; English.
 CC The present sequence represents an autoimmune regulator-1 (AIRE-1)
 CC protein. The AIRE polypeptides and polynucleotides can be used
 CC in methods for the diagnosis and treatment of diseases related
 CC to immune maturation and regulation of immune response towards
 CC self and nonself. They can be used particularly in the diagnosis
 CC and treatment of autoimmune polyendocrinopathy candidiasis
 CC ectodermal dystrophy (APECED) (also known as autoimmune
 CC polyglandular syndrome type 1 (APS I)).
 XX
 SQ Sequence 545 AA;
 Query Match 100.0%; Score 2902; DB 20; Length 545;
 Best Local Similarity 100.0%; Pred. No. 1.5e-211;
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATDALRLRLRLHRTETIAVAVDSAPFLHLALADHDVPEDEKFOETLHKEEGCQPAFH 60
 DB 1 MATDAALRLRLRLHRTETIAVAVDSAPFLHLALADHDVPEDEKFOETLHKEEGCQPAFH 60
 QY 61 ALLSWLLTQDSTAILDFWRVLFKDYNLERYGRLQPIIDSPKDVLSQPRKGRPPAVPK 120
 DB 61 ALLSWLLTQDSTAILDFWRVLFKDYNLERYGRLQPIIDSPKDVLSQPRKGRPPAVPK 120
 QY 121 ALVPPRLPTKRKASEEARAAPALTPRGTAASPSQKAKPKPKPSSAEOQRLPLNG 180
 DB 121 ALVPPRLPTKRKASEEARAAPALTPRGTAASPSQKAKPKPKPSSAEOQRLPLNG 180


```

OY 181 IOTMSAVORAVAMSSGDVPGARGAVEGILLIOVFESGSKCICQVGGERTPSKFEFDSG 240
    |||||||
DB 181 IOTMSAVORAVAMSSGDVPGARGAVEGILLIOVFESGSKCICQVGGERTPSKFEFDSG 240
OY 241 SGKNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECA 300
    |||||||
DB 241 SGKNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECA 300
OY 301 VCRDGGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEP 360
    |||||||
DB 301 VCRDGGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEP 360
OY 361 PVFPLPGLRSAGEEVRPPGEPPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCY 420
    |||||||
DB 361 PVFPLPGLRSAGEEVRPPGEPPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCY 420
OY 421 GPEGQONLAPGARGCGVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGSDVT 480
    |||||||
DB 421 GPEGQONLAPGARGCGVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGSDVT 480
OY 481 PAPVEGVLAAPSPARLAPGPAKDDTASHEPALHRDLESLLSEHTFDGILLQMATQSMARPA 540
    |||||||
DB 481 PAPVEGVLAAPSPARLAPGPAKDDTASHEPALHRDLESLLSEHTFDGILLQMATQSMARPA 540
OY 541 APFPS 545
    |||||
DB 541 APFPS 545

RESULT 3
AA06074
ID AA06074 standard; Protein: 552 AA.
AC
AA06074:
XX
XX 16-AUG-1999 (first entry)
DE
XX
XX Mouse AIRE protein.
KW Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
KM APECED; autoimmune polyglandular disease type 1; APGD; AIRE;
KM transcription factor; autoimmune disease; diagnosis; gene therapy;
KM mouse; animal model.
OS
XX
XX Mus sp.
XX
XX WO9918197-A2.
XX
XX 15-APR-1999.
XX
XX 02-OCT-1998; 98WO-EP06294.
XX
XX 12-NOV-1997; 97EP-0119810.
XX 02-OCT-1997; 97EP-0117154.
XX 08-OCT-1997; 97EP-0117398.
XX
XX (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
XX (NAPU-) NAT PUBLIC HEALTH INST.
XX
XX Aallonen J, Bjorses P, Horelli-Kuitunen N, Lehnach H;
XX Palotie A, Peltonen L, Perheentupa J, Yaspo M;
XX WPI; 1999-287735/24.
XX N-PSDB; AAX58606.
XX
XX
XX New polypeptide which co-segregates in mutated form - with
XX PT autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
XX PS Example 20; Fig 14A-C; 77pp; English.
XX
XX This is the amino acid sequence of murine AIRE, as deduced from
XX CC an isolated cDNA clone (see AAX58606). Murine AIRE is the homologue

```

```

CC of the human AIRE (or autoimmune polyglandular disease type 1,
CC APECED) polypeptide (see AA06073). The overall identity between the
CC mouse and human AIRE proteins is 72.37%. Human mutated APGD1
CC co-segregates with autoimmune polyendocrinopathy candidiasis
CC ectodermal dystrophy (APECED). The murine homologue may be used to
CC develop an animal model for APECED, to examine the events that lead
CC to the development of APECED and possibly to develop agents for
CC preventing and/or treating this autoimmune disease.
XX
XX Sequence 552 AA:
XX
XX Query Match 69.1%; Score 2005; DB 20; Length 552;
XX Best Local Similarity 71.8%; Pred. No. 1.5e-143;
XX Matches 395; Conservative 37; Mismatches 108; Indels 10; Gaps 6;
OY 4 DAALRLRLRLRTEIAVAVDSAFPLHLADHDVVPBEKFGQTLILKEEGCPQAFHALL 63
    |||||||
DB 5 DGMRLRLRLRLRTEIAVAVDSAFPLHLADHDVVPBEKFGQTLILKEEGCPQAFHALL 64
OY 64 SWLLTODSTATILDPMWRLFKDYNLERYGRLOPLDLSFKDVLDSOPRKGRKPPAVKALV 123
    |||||||
DB 65 SWLLTODSTATILDPMWRLFKDYNLERYGRLOPLDLSFKDVLDSOPRKGRKPPAVKALV 124
OY 124 PPPRLPTRKRKASEARAAAPALPRTGASPGSQLKAPPKPESSAEQRLPLNGIQT 183
    |||||||
DB 125 LPPRPPTKRKALEBRATPPATLASKSVSPSHLTKPKPKPDGNTLESQHLPLNGIQT 184
OY 184 MSASVORAVAMSSGDVPGARGAVEGILLIOVFESGSKCICQVGGERTPSKFEFDSG 242
    |||||||
DB 185 MAASVORAVAVASGDVPGARGAVEGILLIOVFESGSKCICQVGGERTPSKFEFDSG 244
OY 243 KNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECAVC 302
    |||||||
DB 245 KNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECAVC 304
OY 303 RDGGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEPV 362
    |||||||
DB 305 HDGGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEPV 364
OY 363 ETPPLPGLRSAGEEVRPPGEPPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCYGP 422
    |||||||
DB 365 ETPPLPGLRSAGEEVRPPGEPPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCYGP 422
OY 423 EGOONLAPGARGCGVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGSDVT 482
    |||||||
DB 423 EGRGPAPARSAKSCVCGDGTDLRCAHCAAFHMRCHFPAGTSRPGTGLRCRSGSDVT 482
OY 483 P-VEGVLAAP-SPARLAPGPAK--DITASHEPALHRDLESLLSEHTFDGILLQMATQSMAR 538
    |||||||
DB 483 PGTGGEAVPTSGPRAPGLAKVGDSDASHDPLVLAHRDLESLLSEHTFDGILLQMATQSMAR 542
OY 539 PAA---PFP 545
    |||||
DB 543 PLAEPTPFSS 552

RESULT 4
AA01713
ID AA01713 standard; Protein: 348 AA.
AC
AA01713:
XX
XX 24-JUN-1999 (first entry)
XX
XX A human autoimmune regulator-2 (AIR-2) protein.
XX
XX Autoimmune regulator-2; AIR-2; Immune maturation; Immune response;
XX KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
XX KW APECED; autoimmune polyglandular syndrome type 1; APS 1.
XX
XX Homo sapiens.
XX
XX WO9915559-A1.

```

```

XX 01-APR-1999.
PD 23-SEP-1998; 98WO-F100749.
XX 23-SEP-1997; 97FI-0003762.
PR (FIIM-) FINNISH IMMUNOTECHNOLOGY LTD.
XX Antanarakis S, Helmo M, Krohn K, Kudoh J, Lalioti M;
PI Peterson P, Scott H, Shimizu N;
XX WPI; 1999-244390/20.
DR N-PSDB; AAX26937.
XX Autoimmune regulator 1 (AIR1) DNA sequence
PS Claim 7; Page 29-30; 59pp; English.
XX The present sequence represents an autoimmune regulator-2 (AIR-2)
CC protein. The AIR polypeptides and polynucleotides can be used
CC in methods for the diagnosis and treatment of diseases related
CC to immune maturation and regulation of immune response towards
CC self and nonself. They can be used particularly in the diagnosis
CC and treatment of autoimmune polyendocrinopathy candidiasis
CC ectodermal dystrophy (APECED) (also known as autoimmune
CC polyglandular syndrome type I (APS I)).
XX Sequence 348 AA;

Query Match 49.1%; Score 1423.5; DB 20; Length 348;
Best Local Similarity 83.5%; Pred. No. 9,7e-100;
Matches 269; Conservative 10; Mismatches 38; Indels 5; Gaps 3;

QY 227 GGEYTPSKFEDSGSGKNKRRSSGPKPLVRAKAGAGAPG---GGEARLGGQGSVPAPL 283
DB 29 GGVCWRPDGKGTGGGRISGPGSMAGQRLGSSGTQRCWCSGCEKEVAL--RRVLHPSPV 87
QY 284 ALPSDPLQKNEDECAVNCRDGELICDCCPRAFHLACSLPRLREIPSGTWRSSCLQA 343
DB 88 CM-GVSCLOKNEDECAVNCRDGELICDCCPRAFHLACSLPRLREIPSGTWRSSCLQA 146
QY 344 TVQEVQRAEPRPQEPVETPLPGLRSAGEEVRGPPGEPPLAGMDTTLVYKHLPAAPSA 403
DB 147 TVQEVQRAEPRPQEPVETPLPGLRSAGEEVRGPPGEPPLAGMDTTLVYKHLPAAPSA 206
QY 404 APPLGIDSSALHPLLCVGPBGQONLAPGARCGVCGDGTDLVLRCTHCAAFHMRCHFPAGT 463
DB 207 APPLGIDSSALHPLLCVGPBGQONLAPGARCGVCGDGTDLVLRCTHCAAFHMRCHFPAGT 266
QY 464 SRPSTGLRCRSCSGDYTPAPEGVLAAPSAPARLAPGPAKDDTASHEPDLHRDLESLSSEH 523
DB 267 SRPSTGLRCRSCSGDYTPAPEGVLAAPSAPARLAPGPAKDDTASHEPDLHRDLESLSSEH 326
QY 524 TPFGILLQMAIQSMARPAAPPS 545
DB 327 TEDGILLQMAIQSMARPAAPPS 348

RESULT 5
AAV01714
ID AAV01714 standard; Protein; 254 AA.
XX
XX AAV01714;
AC
XX
XX 24-JUN-1999 (first entry)
DT
XX
XX A human autoimmune regulator-3 (AIR-3) protein.
DE
XX Autoimmune regulator-3; AIR-3; immune maturation; immune response;
KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
APCED; autoimmune polyglandular syndrome type I; APS I.
XX

```

```

OS Homo sapiens.
XX
XX W09915559-A1.
PN
XX 01-APR-1999.
PD
XX 23-SEP-1998; 98WO-F100749.
PF 23-SEP-1997; 97FI-0003762.
XX (FIIM-) FINNISH IMMUNOTECHNOLOGY LTD.
PR 23-SEP-1997; 97FI-0003762.
XX Antanarakis S, Helmo M, Krohn K, Kudoh J, Lalioti M;
PI Peterson P, Scott H, Shimizu N;
XX WPI; 1999-244390/20.
DR N-PSDB; AAX26938.
XX Autoimmune regulator 1 (AIR1) DNA sequence
PS Claim 7; Page 32-33; 59pp; English.
XX The present sequence represents an autoimmune regulator-3 (AIR-3)
CC protein. The AIR polypeptides and polynucleotides can be used
CC in methods for the diagnosis and treatment of diseases related
CC to immune maturation and regulation of immune response towards
CC self and nonself. They can be used particularly in the diagnosis
CC and treatment of autoimmune polyendocrinopathy candidiasis
CC ectodermal dystrophy (APECED) (also known as autoimmune
CC polyglandular syndrome type I (APS I)).
XX Sequence 254 AA;

Query Match 17.4%; Score 503.5; DB 20; Length 254;
Best Local Similarity 43.8%; Pred. No. 3,5e-30;
Matches 126; Conservative 17; Mismatches 76; Indels 69; Gaps 11;

QY 227 GGEYTPSKFEDSGSGKNKRRSSGPKPLVRAKAGAGAPG---GGEARLGGQGSVPAPL 283
DB 29 GGVCWRPDGKGTGGGRISGPGSMAGQRLGSSGTQRCWCSGCEKEVAL--RRVLHPSPV 87
QY 284 ALPSDPLQKNEDECAVNCRDGELICDCCPRAFHLACSLPRLREIPSGTWRSSCLQA 343
DB 88 CM-GVSCLOKNEDECAVNCRDGELICDCCPRAFHLACSLPRLREIPSGTWRSSCLQA 146
QY 344 TVQEVQRAEPRPQEPVETPLPGLRSAGEEVRGPPGEPPLAGMDTTLVYKHLPAAPSA 403
DB 147 TVQEVQRAEPRPQEPVETPLPGLRSAGEEVRGPPGEPPLAGMDTTLVYKHLPAAPSA 192
QY 404 APPLGIDSSALHPLLCVGPBGQONLAPGARCGVCGDGTDLVLRCTHCAAFHMRCHFPAGT 463
DB 193 -----KCVW-----VLRVSR-----MLVAVAGC 212
QY 464 SRPSTGLRCRSCSGDYTPAPEGVLAAPSAPARLAPGPAKDDTASHEPDL 510
DB 213 AE-----MKRTCCGVLTAPLPSTGATSQPA--PPGPR--ACAADPA 251

RESULT 6
AAR99534
ID AAR99534 standard; Protein; 1911 AA.
XX
XX AAR99534;
AC
XX
XX 30-OCT-1996 (first entry)
DT
XX
XX Dermatomyositis specific autoantigen, Mi-2.
DE
XX Mi-2; autoantigen; collagen disease; chromosome 12; 12p13;
KW helicase; dermatomyositis; diagnosis.
XX
XX Homo sapiens.
OS

```

Key	Location/Qualifiers
FT	53..73
FT	/label= Region_a
FT	/note= "contains 3 potential core target motifs"
FT	113..133
FT	/label= Region_b
FT	/note= "contains 4 potential core target motifs"
FT	133..143
FT	/note= "possible electrostatic interaction with chromatin of histones"
FT	257..287
FT	/label= Region_c
FT	/note= "contains 3 potential core target motifs"
FT	747..758
FT	/note= "Helicase-specific motif I"
FT	782..793
FT	/note= "Helicase-specific motif IA"
FT	869..877
FT	/note= "Helicase-specific motif II"
FT	897..911
FT	/note= "Helicase-specific motif IIR"
FT	943..959
FT	/label= Region_d
FT	/note= "contains 1 potential core target motif"
FT	949..960
FT	/note= "Helicase-specific motif IV"
FT	1121..1144
FT	/note= "Helicase-specific motif V"
FT	1149..1183
FT	/note= "Helicase-specific motif VI"
XX	DEL9509279-C1.
XX	15-MAY-1996.
XX	15-MAR-1995; 95DE-1009279.
XX	15-MAR-1995; 95DE-1009279.
XX	(PRIV-) PRIVATEs INST IMMUNOLOGIE & MOLEKULARGEN.
PA	Renz M, Seelig HP;
PI	WPI: 1996-240280/25.
DR	N-PSDB: AAT32301.
XX	DNA encoding dermatomyositis specific auto:antigen - useful for differential diagnosis and treatment of dermatomyositis
PT	Claim 1; Fig 2; 20pp; German.
XX	The present sequence is that of a 218 kD dermatomyositis specific auto-antigen, designated MI-2. The sequence numbering given in the specification starts at amino acid 2, i.e. the first Met residue is omitted. The protein is hydrophilic, acidic and protruding regions of the protein are characteristic of helicases. The gene corresponding to the cDNA (AAT32301) encoding MI-2, was localised to chromosome 12 (12p13). The DNA can be used for the recombinant production of MI-2 which is used for, e.g. the differential diagnosis of collagen diseases, esp. dermatomyositis, e.g. by immunoassay or Western blotting.
XX	Sequence 1911 AA;
XX	Query Match 8.3%; Score 241; DB 17; Length 1911;
XX	Best Local Similarity 23.1%; Pred. No. 3.3e-09;
XX	Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14
OY	129 PTKRAASEARAAALPTPGTASBSCLOKKPKPKPPSSABQORLPLGNGIQTWSASV 188
DB	216 PFKGSSGASVAANAANAANVWESMTATDEV-APPPEVVPPIKAKTKGKGFNA----- 269
OY	189 CRAVAMSSODVVGACAVGECILIQVFEESG-----SKKCIQVGGEYTPSKPED----- 238

Db	270	-RRKPGSPRRVPAPKPKPKKVAAPLKIKLGGFGSKRRRSSSEDDDLVESDFDASINSY	328
Qy	239	SSSGKNKANSSSGPKPLVRAKGAQAAPGSGEARLGGQGVAPALALPSDFOLHOKNE	298
Db	329	SVSDGTSRSSRSRKKLRITTK-----KKKKGE-EVYAVDGETD-----HQDY	371
Qy	299	CAVCRNGGELLICDDGPRAFHLACLSPPLREIPSGTGRSSCLQATVQEVQPRAEPRRQ	358
Db	372	CEVCQGGGELLICDDTPRAAHMVCLOPDMKAPRGKSCPCRK--EGIQWEAKED--	425
Qy	359	EPPVETPLPPGLSRASGEVGPGEPLAGADDTLVYHHLPAAPSAAPLPGIDSSALHPL	418
Db	426	-----NSEEELIEEYGGDLSEEDD-----HIMEF-----	450
Qy	419	CVGPEGQNLAPGARCGVCGDGTIVLRCTHCAAFHWRCHFPACTSRPGTLRCRSCGD	478
Db	451	-----CRVCKDGGELLICDDCPSSYHICLNPPLPEIPNGEMLCPRTCP	495
Qy	479	VTFPAPVEGL-----APSPARL-ARGPARDQTRASHPALHRDLSELSLSEHTDGIQW	531
Db	496	ALRGKVKOKILIMKMGQPSPTPVPRPDADPNTPSPRLRGPRPERQFVVK---WQMSYW	552
RESULT 7			
AAAM39288			
ID	AAAM39288	standard; Protein; 1912 AA.	
XX	AAAM39288;		
AC			
XX			
DT	22-OCT-2001	(first entry)	
DE			
XX	Human polypeptide SEQ ID NO 2433.		
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX			
OS	Homo sapiens.		
XX			
FN	WO200153312-A1.		
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Dirmannac RT;		
XX			
DR	WPI: 2001-442253/47.		
XX	N-PSDB; AAI58444.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Example 4: SEQ ID NO 2433; 10078bp; English.		
XX			
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and		
CC	the encoded polypeptides (AAAM38642-AAAM42213) with nootropic,		

Db 518 ALKGVQKILMKMGQSPSPVPRPPDADPTPSPKPLEGRERFVVK---WQMSYW 574

RESULT 9

ABB07636 ID ABB07636 standard; Protein: 689 AA.

AC ABB07636;

XX 20-MAY-2002 (first entry)

XX Human speckled 110 (Sp110) polypeptide.

XX

XX Sp110: speckled 110: nuclear body; gene transcription; dimerisation;

XX nuclear hormone receptor; primary biliary cirrhosis; PBC; human;

XX cell differentiation; cellular defence; cytosolic; anorectic;

XX antiinflammatory.

XX Homo sapiens.

XX

XX Key Location/Qualifiers

XX Domain 6..109

XX /note= "Sp100-like domain"

XX Domain 454..532

XX /note= "SAND domain"

XX Domain 537..577

XX /note= "plant homeobox domain"

XX Domain 606..674

XX /note= "bromodomain"

XX

XX WO200208383-A2.

XX

XX 31-JAN-2002.

XX

XX 24-JUL-2001: 2001WO-US23248.

XX

XX 24-JUL-2000: 2000US-220305P.

XX

XX (GEHO) GEN HOSPITAL CORP.

XX

XX Bloch DB, Bloch KD;

XX WPI: 2002-227042/28.

XX DR N-PSDB: ABA95062.

XX

XX Novel substantially pure polypeptide component of the nuclear body,

XX speckled 110, useful in screening methods, in clinical diagnostic

XX methods, and for treating myeloid malignancies, inflammation and

XX obesity -

XX

XX Claim 3: Fig 2: 55pp; English.

XX

XX The invention relates to cloning and characterisation of a cDNA encoding

XX Sp110 (speckled 110), a novel 110 kDa polypeptide component of the

XX nuclear body. The Sp110 polypeptide comprises a Sp100-like domain, a

XX SAND domain, a plant homeobox domain, and a bromodomain. Sp110 functions

XX as an activator of gene transcription and serves as a nuclear hormone

XX receptor co-activator. Sp110 is useful in screening for identifying a

XX compound that modulates (i) Sp110 dimerisation; (ii) Sp110 binding to a

XX nuclear hormone receptor; (iii) binding of an Sp110 dimer to an Sp110-

XX binding nucleotide sequence and in a screening method for identifying a

XX polypeptide that dimerises with Sp110 to form a constitutively active or

XX hyperactive or inactive heterodimer. It is useful for diagnosing primary

XX biliary cirrhosis (PBC) and is also useful for producing Sp110-specific

XX antibodies, for inhibiting viral replication and facilitating

XX differentiation of cells, e.g. myeloid cells, and activation of cells

XX involved in host defence, to treat myeloid malignancies, to enhance

XX cellular defence mechanisms, to treat inflammation, to achieve

XX alteration in lipid profiles, to block estrogen receptors in treatment

XX of estrogen responsive tumours, and for treating obesity. The present

XX sequence represents the human Sp110 polypeptide.

SQ Sequence 689 AA:

Query Match 8.2%; Score 238; DB 23; Length 689;

Best Local Similarity 19.7%; Pred. No. 1,6e-09;

Matches 120; Conservative 44; Mismatches 137; Indels 308; Gaps 19;

OY 2 AYDAALRLRLRLHRTETIAVAVDSAFPLHLALADHDVPEDEKFOETLHKEKEGC----- 55

DB 7 AMEEALFOHFMHQKLGIAVAIHKPFPEEGILDSITTKRYMESL-----EACNLLIPV 61

OY 56 PQAFHALLSWLTLQDSTALIDFWRVLFKDYVLEFRGLQPLIDSEPKDVDLSQPRKGRKP 115

DB 62 SRVYHNLITQL---ERTFNLSLVTFLFSQILNREYPNLVTYIRSP-KRVGASVYERQSRDT 117

OY 116 PAV-----PRALVPP----- 125

DB 118 PILLLEAPTLGAESSLHTPLALPPQPPQPCSPCAPVSEPTSSQOSDEILSESPSS 177

OY 126 -PRLP-----TKRKASEARAAAPALT----- 147

DB 178 DPVLPPLPALIQEGRSTSVTNDKLTSKMAEEDSEMPSLISTVQVYASDNLIPQIRKED 237

OY 148 ----- 147

DB 238 PQEMPHSPGSMPEIRONSPEPNDEPEQEVSTSPDCKKKKRCIMSTPKRHKKSL 297

OY 148 PRGTASP--GSQLKAK-----PKKPPESSA----- 170

DB 298 PRGTASSRHGIQKRLKRVQYPOKKDDSTCNSTVETRAQKARTECARKSRSEIIDGTSE 357

OY 171 -----EQRRLPL-----GNGIO-----TMSASVORA-- 191

DB 358 MNEGKRSQKTPSTPRRYTQGAASPGHGIEKLVQVVDKYQQRKDDSTWNSVMMRVQKART 417

OY 192 -----VAMSSGDVPCARGAVEGILIQ 212

DB 418 KCAKRSKSEKKEKEDICSSSKRRFOKNIHRRGKPKSDVDFHCSKLPVTCGEAKGILYK 477

OY 213 QVESGSGSKKCIQ--VGGEFTYPSKPEDSGSKNKRASSGPRPLVRAKGAQGAAPGGEA 271

DB 478 KMKRHGSSVCKICRNEDGTWLPNFEFEVEGKGRN---AKNWKNIKCEG-----M 523

OY 272 RLGGGGVPAVLAIPSDPOLHOKNEDECAVCDGGLICCDGCPRAFHILACLSPPLEIRP 331

DB 524 TLGE-----LKRKNSDECVCCGGGLCCGTCPRVFHECHILPV-EAK 568

OY 332 SGTWRCSSC 340

DB 569 RMLWSCFRC 577

RESULT 10

AAR22280 ID AAR22280 standard; Protein: 351 AA.

XX

XX AAR22280;

XX

XX 27-JUL-1992 (first entry)

XX

XX M1-2 antigen clone L1 encoded protein.

XX

XX Myositis-specific antigen; polymyositis scleroderma overlap;

XX syndrome; dermatomyositis; autoantibodies; immunodiagnostic assay.

XX

XX Homo sapiens.

XX OS

XX WO9204472-A.

XX PN

XX 19-MAR-1992.

XX PD

XX 05-SEP-1991: 91WO-US06418.

XX PF

XX 07-SEP-1990: 90US-0579023.

XX PR

YY	RESULT 12
XX	AAM93902
XX	ID AAM93902 standard; Protein; 1050 AA.
XX	AC AAM93902;
XX	DT 06-NOV-2001 (first entry)
XX	DE Human polypeptide, SEQ ID NO: 4045.
XX	KW Human; full length cDNA; cDNA synthesis; oligo-capping.
OS	Homo sapiens.
PX	EP1130094-A2.
PN	05-SEP-2001.
PD	07-JUL-2000; 2000EP-0114089.
PF	08-JUL-1999; 99JP-0194486.
PR	11-JAN-2000; 2000JP-0118774.
PR	02-MAY-2000; 2000JP-0183765.
XX	(HELI-) HELIX RES INST.
PA	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI	Makematsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX	WPI: 2001-524255/58.
DR	N-PSTDB; AAK94864.
PT	830 Primers useful for synthesizing full length cDNA clones and their
PP	use in genetic manipulation -
PS	Claim 8; SEQ ID NO 4045; 1380pp + sequence listing; English.
XX	The invention relates to primers for synthesizing full length cDNA
CC	clones. 830 cDNA molecules encoding a human protein have been
CC	isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC	molecules have been determined. Primers for synthesizing the full length
CC	cDNA are useful for clarifying the function of the protein encoded by
CC	the cDNA. The full length clones were obtained by construction of full
CC	length enriched cDNA libraries that were synthesised by the oligo-capping
CC	method. The primers enable the production of the full length cDNA easily
CC	without any special methods. The present sequence is a polypeptide
CC	encoded by a full length human cDNA of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in CD-ROM format directly from ERO.
XX	Sequence 1050 AA;
SO	
QY	Query Match 7.8%; Score 227; DB 22; Length 1050;
	Best Local Similarity 24.6%; Pred. No. 1.8e+08;
	Matches 77; Conservative 46; Mismatches 104; Indels 86; Gaps 10
DY	71 STAILDFMVLKLDNYLERYGRLOPILDSFPRKDVISOPRKGGKKPAVPRAIYPRLPT 130
DB	631 STIMD--NIYKRDNIID-HGQRPSPSNTVOSSPVNSPGLAGVTMTSVHPTRSPS 687
DY	131 KRKASEEAAAPALTPRGTA SPG-----SOLKAKPKR-----PESSA 170
DB	688 --ASSVGSKSGSSGSSKPGADSTRHVPVMLEPIRIKQENSGPENNPDPVTVKQESD 745
DY	171 EQORLPGLGIDTMSASVQRAVAMS-----GVPPGARCAVEIIILIQVRESGSKR 222
DB	746 ESRQNOMATYPPSIITSLILNSSOSTSEFYVLRSDAPPSTGQPGL-----792
DY	223 CIOVGGEFTYPERKFEDSGSKNKARSSGPKPLVRAKAGQAAPAGGGEARLGOOGSVAP 282
DB	793 -----HDNMSGKSEMIDPQGKSP L-----HNGEIR-----819
DY	283 IALPSPDLQHKNNEDECAVCARDGEGELICDDGCPBRAHACLSPPLREITSIGWRSCCQ 342

Db	820 --KEDP-----NEMWCAYCNGSGELLCCEKCPKFVHLSCHVPYTLTFPSGEGICTFCD	872
Qy	343 ATVEGVOPRAEPP 355 : : Db 873 LSKPEVEYDCDAP 885	
 RESULT 13		
ABBO7637	ABB07637 standard; Protein; 753 AA.	
XX	ABB07637;	
XX	20-MAR-2002 (first entry)	
DE	Human speckled 140 (Sp140) polypeptide.	
XX	Sp110; speckled 110; nuclear body; gene transcription; dimerisation;	
KW	nuclear hormone receptor; primary biliary cirrhosis; PBC; human; Sp140;	
KM	cell differentiation; cellular defence; cytostatic; anorectic;	
KX	antiinflammatory.	
XX		
OS	Homo sapiens.	
PN	WO200208383-A2.	
XX		
PD	31-JAN-2002.	
PF	24-JUL-2001; 2001WO-US23248.	
XX		
PR	24-JUL-2000; 2000US-220305P.	
XX		
PA	(GEHO) GEN HOSPITAL CORP.	
XX		
PI	Bloch DB, Bloch KD;	
DR	WPI, 2002-227042/28.	
PT		
PT	Novel substantially pure polypeptide component of the nuclear body,	
PT	speckled 110, useful in screening methods, in clinical diagnostic	
PT	methods, and for treating myeloid malignancies, inflammation and	
PT	obesity -	
XX		
PS	Disclosure: Fig 2; 55pp: English.	
XX		
CC	The invention relates to cloning and characterisation of a cDNA encoding	
CC	Sp110 (speckled 110), a novel 110 kDa polypeptide component of the	
CC	nuclear body. The Sp110 polypeptide comprises a Sp100-like domain, a	
CC	SAND domain, a plant homeobox domain, and a bromodomain. Sp110 functions	
CC	as an activator of gene transcription and serves as a nuclear hormone	
CC	receptor co-activator. Sp110 is useful in screening for identifying a	
CC	compound that modulates (i) Sp110 dimerisation; (ii) Sp110 binding to a	
CC	nuclear hormone receptor; (iii) binding of an Sp110 dimer to an Sp110-	
CC	binding nucleotide sequence and in a screening method for identifying a	
CC	polypeptide that dimerises with Sp110 to form a constitutively active or	
CC	hyperactive or inactive heterodimer. It is useful for diagnosing primary	
CC	biliary cirrhosis (PBC) and is also useful for producing Sp110-specific	
CC	antibodies, for inhibiting viral replication and facilitating	
CC	differentiation of cells, e.g. myeloid cells, and activation of cells	
CC	involved in host defence, to treat myeloid malignancies, to enhance	
CC	cellular defence mechanisms, to treat inflammation, to achieve	
CC	alteration in lipid profiles, to block estrogen receptors in treatment	
CC	estrogen responsive tumours, and for treating obesity. The present	
CC	sequence represents the human Sp140 polypeptide, which recruits Sp110	
CC	to the nuclear body.	
XX		
SQ	Sequence 753 AA;	
Query Match	7.6%; Score 220.5; DB 23; Length 753;	
Best Local Similarity	20.5%; Pred. No. 3.7e-08;	
Matches 128; Conservative	53; Mismatches 152; Indels 291; Gaps 24;	

```

QY 7 LRRLRLHRTETAVAVDSAFPLLHALADHDVPE---DKFOETLHLKEKSGCPQAFHALL 63
DB 42 IRRFRRENVEIASATIRPFPLMGLRDSFISEQWEHFOEA--FRNLVPTVRVMVCVL 99
QY 64 SWLLQDSTAILDFWVFLKRDYLERGRLOPLIDSF----- 100
DB 100 SEL---EKTFGWSHLEALFSRLNLMAVYPDLEIYRSFQVNCYEHSPLOMNNVNDLEDRPR 156
QY 101 -----PKVDLSOPR-----GKRKPAPVAKA-----L 122
DB 157 LIPYKQENSNACHEMDDTAVQEAALSSSPKCPGSSSECEQALPKAGGDAEDAPSL 216
QY 123 VP-----PPRLP-----TKRKASE----- 136
DB 217 LPVSKLAIQIDEGESEEMPKLLPYDTEFTDLKTPQVYNESGPEKGLCLLPGEGERSD 276
QY 137 -----EARAAPALTPRGTS-----PSQLKA-- 160
DB 277 DCEMCDGEERQEAASSLARSGVSSELENHPNNEGESEELASSLLYDNVPGAEQSAVE 336
QY 161 -----KPKKPSSAE-----QORLPNG----- 180
DB 337 NEKSCVWCSEEPVGSPEARTESDQACGTMDTVDIANNSTLGKPKRKRKRKGHWSRM 396
QY 181 -IQTMSASVQRAVAMSSGDV-----PGAR----- 203
DB 397 RMRQKNSQNDNSKADQVVSSEKKANVNLKLSKIRGKRKGKPGTRFTQSDRAAQKV 456
QY 204 -----GAVGILLIQVFESGSKKCIQV--GGEFTYPSKED 238
DB 457 RSRASRKHKDETVDKAPLLPVTCGVGKGLHKRKLQOGLIVKCIQTEDGKWFTEPEFI 516
QY 239 SGSGKNARSSSGPRPLVRAKAGAGAPGGEAR--LGOOGSVAPPLAL----- 285
DB 517 KG---GHARSKNW--RLSVRC-----GGMPLRLMLMENGFLPDPPIRYRKKKRLKSO 564
QY 286 ---PSDQLHOKNDECAVCRDGGELICDGCPRAFHLACLSPLREIPSGTWRCSSC-- 340
DB 565 NNSVDPCM--RNLDECEVCRDGGELFCDDCSRVFHEHCHIPPV--EAERTPMNCIFCRM 621
QY 341 ----LQATVOEVQRAEPRPQE 359
DB 622 KESGSOQCCQESSEVLERQKCPPE 645

RESULT 14
AAW57747
ID AAW57747 standard; Protein; 753 AA.
AC AAW57747;
DT 17-SEP-1998 (first entry)
DE Leukocyte specific protein, Spi140.
KW Spi140; leukocyte specific protein; gene transcription regulator; therapy;
KW autoimmune disease; viral infection; cancer.
OS Homo sapiens.
PN W09814569-A1.
PD 09-APR-1998.
PE 02-OCT-1997; 97WO-US17715.
PR 02-OCT-1996; 96US-0027347.
PA (BLOC/) BLOCH D B.
PI (BLOC/) BLOCH K D.
XX Bloch DB, Bloch KD;

```

```

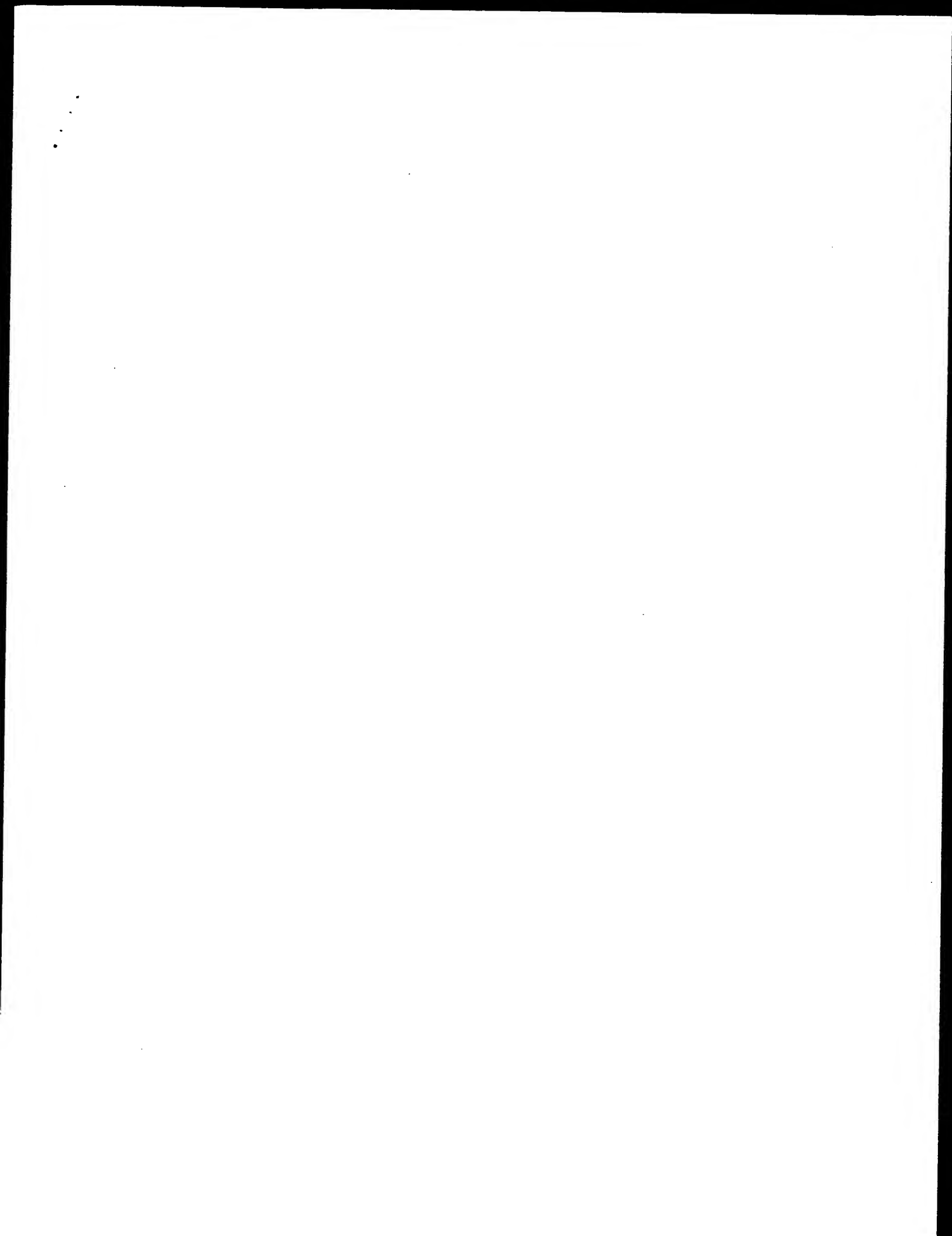
DR WPI: 1998-286419/25.
DR N-PSDB: AAW24559.
XX
PT New isolated gene transcription regulator, Spi140 - used to develop
PT products for the diagnosis and treatment of auto-immune diseases,
PT viral infections or cancers
XX
PS Claim 11; Page 54-57; 81pp; English.
XX
CC This sequence represents the leukocyte specific protein, Spi140 of the
CC invention. The Spi140 polypeptides act as gene transcription regulators.
CC They can be used to develop products for use in the diagnosis and
CC treatment of autoimmune diseases such as primary biliary cirrhosis,
CC rheumatoid arthritis, systemic lupus erythematosus, Sjogren's syndrome,
CC scleroderma and multiple sclerosis, viral diseases including those caused
CC by herpes simplex virus, cytomegalovirus, HIV, hepatitis virus, human
CC T-cell leukemia virus-1 (HTLV-1) and adenovirus, and cancers including
CC breast, ovary, prostate, bone, liver, pancreas or spleen, sarcomas and
CC melanomas.
XX
SQ Sequence 753 AA;

Query Match 9 LRRLRLHRTETAVAVDSAFPLLHALADHDVPE---DKFOETLHLKEKSGCPQAFHALLSW 65
Best Local Similarity 20.78; Score 213.5; DB 19; Length 753;
Matches 129; Conservative 54; Mismatches 148; Indels 291; Gaps 26;

QY 9 LRRLRLHRTETAVAVDSAFPLLHALADHDVPE---DKFOETLHLKEKSGCPQAFHALLSW 65
DB 44 RFRRENVEIASATIRPFPLMGLRDSFISEQWEHFOEA--FRNLVPTVRVMVCVLSE 101
QY 66 LQDSTAILDFWVFLKRDYLERGRLOPLIDSF-----PKDV 104
DB 102 L---EKTFGWSHLEALFSRLNLMAVYPDLEIYRSFQVNCYEHSPLOMNNVNDLEDRPRL 158
QY 105 -----DLSDPRK-----GKRKP-----ATPKA-----LVP 124
DB 159 PYKQENSNACHEMDDTAVQEAALSSSPKCPGSSSECEQALPKAGGDAEDAPSLP 218
QY 125 -----PPRLP-----TKRKASE----- 136
DB 219 VSKLAIQIDEGESEEMPKLLPYDTEFTDLKTPQVYNESGPEKGLCLLPGEGERSDC 278
QY 137 -----EARAAPALTPRGTS-----PSQLKA-- 160
DB 279 SEMCDGEERQEAASSLARSGVSSELENHPNNEGESEELASSLLYDNVPGAEQSAVENE 338
QY 161 -----KPKKPSSAE-----QORLPNG-----I 181
DB 339 KCSVCWCSEEPVGSPEARTESDQACGTMDTVDIANNSTLGKPKRKRKRKGHWSRM 398
QY 182 QTMASVQRAVAMSSGDV-----PGAR----- 203
DB 399 RQKNSQNDNSKADQVVSSEKKANVNLKLSKIRGKRKGKPGTRFTQSDRAAQKVRS 458
QY 204 -----GAVGILLIQVFESGSKKCIQV--GGEFTYPSKEDSG 240
DB 459 RASRKHKDETVDKAPLLPVTCGVGKGLHKRKLQOGLIVKCIQTEDGKWFTEPEIKG 518
QY 241 SGKNARSSSGPRPLVRAKAGAGAPGGEAR--LGOOGSVAPPLAL----- 285
DB 519 ---GHARSKNW--RLSVRC-----GGMPLRLMLMENGFLPDPPIRYRKKKRLKSONN 566
QY 286 -PSDQLHOKNDECAVCRDGGELICDGCPRAFHLACLSPLREIPSGTWRCSSC-- 340
DB 567 SSVDCM--RNLDECEVCRDGGELFCDDCSRVFHEHCHIPPV--EAERTPMNCIFCRMKE 623
QY 341 ----LQATVOEVQRAEPRPQE 359
DB 624 SPGSOQCCQESSEVLERQKCPPE 645

RESULT 15

```

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:54:02 : Search time 17 Seconds
(without alignments)
1477.658 Million cell updates/sec

Title: US-09-509-595b-2

Perfect score: 2902

Sequence: 1 MATDALARLRLRLHREINAV.....DGLIQAQISQMARPAPEPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PTCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Print. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	5.8	1134	9	US-09-836-392-16
2	168	5.8	1237	10	US-09-862-027-78
3	166	5.7	549	10	US-09-764-864-1254
4	165	5.7	167	9	US-10-005-057A-31
5	159.5	5.5	1023	9	US-09-893-519A-14
6	156	5.4	432	10	US-09-764-864-1184
7	155	5.3	1400	10	US-09-764-176-7
8	153	5.3	1207	9	US-10-189-971-20
9	151.5	5.2	2462	9	US-09-819-104A-5
10	149	5.1	601	10	US-09-925-301-844
11	148.5	5.1	380	10	US-09-764-176-3
12	147.5	5.1	538	9	US-09-976-740-43
13	147.5	5.1	538	12	US-10-023-529-43
14	147.5	5.0	538	12	US-10-023-523-43
15	145.5	5.0	1251	9	US-10-189-971-16
16	143	4.9	527	9	US-09-854-133-216
17	143	4.9	527	10	US-09-738-973-216
18	142.5	4.9	510	10	US-09-866-562-56
19	140.5	4.8	1274	9	US-10-020-215-2

20	140	4.8	2507	9	US-09-819-104A-2	Sequence 2, Appli
21	139.5	4.8	1056	9	US-10-161-510-10	Sequence 10, Appl
22	139	4.8	550	9	US-09-976-740-47	Sequence 47, Appl
23	139	4.8	550	12	US-10-023-529-47	Sequence 47, Appl
24	139	4.8	550	12	US-10-023-523-47	Sequence 47, Appl
25	136	4.7	522	9	US-09-764-868-1138	Sequence 1138, Ap
26	135.5	4.7	2441	12	US-10-109-886-8	Sequence 8, Appl
27	135	4.7	524	9	US-09-764-868-761	Sequence 761, App
28	135	4.7	846	10	US-09-858-664A-3	Sequence 3, Appl1
29	134.5	4.6	693	9	US-10-029-217A-4	Sequence 4, Appl1
30	134	4.6	1192	9	US-10-189-971-18	Sequence 18, Appl
31	134	4.6	1342	9	US-10-189-971-24	Sequence 24, Appl
32	134	4.6	1477	9	US-10-189-971-8	Sequence 8, Appl1
33	134	4.6	1512	9	US-10-189-971-10	Sequence 10, Appl
34	134	4.6	1535	9	US-10-189-971-14	Sequence 14, Appl
35	134	4.6	1570	9	US-10-189-971-12	Sequence 12, Appl
36	134	4.6	1593	9	US-10-189-971-4	Sequence 4, Appl1
37	134	4.6	1628	9	US-10-189-971-2	Sequence 2, Appl1
38	133.5	4.6	1002	9	US-09-988-117-3	Sequence 3, Appl1
39	133.5	4.6	1002	10	US-09-812-471-3	Sequence 3, Appl1
40	133.5	4.6	1002	10	US-09-812-633-3	Sequence 3, Appl1
41	133	4.6	1317	10	US-09-963-896-7	Sequence 7, Appl1
42	131.5	4.5	420	10	US-09-764-864-1084	Sequence 1084, Ap
43	131.5	4.5	1690	10	US-09-788-043C-5	Sequence 5, Appl1
44	131	4.5	4019	9	US-09-854-133-425	Sequence 425, App
45	131	4.5	4019	10	US-09-738-973-425	Sequence 425, App

ALIGNMENTS

RESULT 1
US-09-836-392-16
Sequence 16, Application US/09836392
Patent No. US20020173458A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Protein Tyrosine Kinase Receptor Polynucleotides, Polypeptides
FILE REFERENCE: PTO20P1
CURRENT APPLICATION NUMBER: US/09/836,392
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: PCT/US00/28066
PRIOR FILING DATE: 2000-10-11
PRIOR APPLICATION NUMBER: 60/159,542
PRIOR FILING DATE: 1999-10-15
PRIOR APPLICATION NUMBER: 60/165,914
PRIOR FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: 60/189,027
PRIOR FILING DATE: 2000-03-14
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 1134
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-392-16
Query Match
Best Local Similarity 26.4%; Pred. No. 0.01;
Matches 111; Conservative 41; Mismatches 180; Indels 88; Gaps 22;
QY 116 PAVPKALVP-----PRLPTKRKASEEARAAAPALTPRGTAAPSGQLAKAPPKRESS 169
DB 767 PLSPLAHTPSPTQASPPPLPGHTVGSHTQSPPAKLHSSPPVVRPKSAEPPRSLK 826
QY 170 AAGQRLPLNGIOTMNASQVAAVAMSSGDV--PGARAVBG-ILIQVFFSGSKKIOY 226
DB 827 RVSASAKLG---ASLSADKKGALKRHSLEVGHPDFRDFGELALHSLASDDETPVEG 883
QY 227 GGEFTYPSKFEEDSGSKNRRASSGPKPLAVRAKGA-----OGAAPGGGEARLGOOGS 278
DB 884 LG---APRGVAVRRLRGROGSPSLIGADPLL--PGASAPPYSSKSKESPGCAEAECTPRAT 939

```
QY 279 VPAPLALPSPD--QLHQKNEDECAVCRDGGELICDGCPRAPHLACTSPLEIPSGTWR 336
D 940 TPGRTLERDVGTCTRHQSVQTE-----DG-----TGMAARAAYAKALSL-PVOEHETG--R 986
QY 337 CSSCLQATVQEVQRAEPPROEPPEVPEPLPGLRSAGEEVRGP--PGEPLA--GMD---- 389
D 987 RSSSGEAGTLPVPIVEPARBGAKAV--VPOPLGADSKGQEPAPLAPSVPEAPRGHERNV 1045
QY 390 -----TTLV-YKHLPAAPSAAPLPGIDSSALHPULCVGPEGQONLAPGACGCGDGT 441
D 1046 LEVEERTTILSGPRSKRASPILSPEP--QTPSLAPAKCSAPSSAVTPVPPA--SLIGSGT 1101
QY 442 DVLKCTHCAAAFHMRCHFPACTSRPGTGLRCRSCSGDVTAPAV--EGVLAPSPARLAPGP 499
D 1102 -----KPOVGLTSCRPAEAVPPAGLTKKGVSSPAP----PGP 1134
```

RESULT 2

```
US-09-862-027-78
; Sequence 78, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1el Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862, 027
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345, 473
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 1237
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-027-78
```

Query Match 5.8%; Score 168; DB 10; Length 1237;
Best Local Similarity 26.4%; Pred. No. 0.011;

Matches 111; Conservative 41; Mismatches 180; Indels 88; Gaps 22;

```
QY 116 PAVPALVP-----PRLPTKKAASEARAAPALTPRGTAAPGSLAKAPPKKRESS 169
D 870 PLSPALHPSPPTQASPPPLPGHTVGSHTTQSPFAKLHSSPVVRRPRKSAEPPRSLK 929
QY 170 AEOQRLPLNGIQTMSAVQRAVAMSSGDV--PGARGAVEG-ILIQVFESSGSKCQIY 226
D 930 RVQSAEKLG---ASLSADKKGALKRHSLEVGHPDFKDFHGLALHSLAESDGETPVEG 986
QY 227 GGEFTTPSKFEDSGGKMKARSSSGPKPLVRAKA-----OGAAPGGGEARLGGQGS 278
D 987 LG---APROVAVRRLGROESPLSLGADPLL--PEGASRPVSSSEKSPGCAEACTPPRAT 1042
QY 279 VPAPLALPSPD--QLHQKNEDECAVCRDGGELICDGCPRAPHLACTSPLEIPSGTWR 336
D 1043 TPGRTLERDVGTCTRHQSVQTE-----DG-----TGMAARAAYAKALSL-PVOEHETG--R 1089
QY 337 CSSCLQATVQEVQRAEPPROEPPEVPEPLPGLRSAGEEVRGP--PGEPLA--GMD---- 389
D 1090 RSSSGEAGTLPVPIVEPARBGAKAV--VPOPLGADSKGQEPAPLAPSVPEAPRGHERNV 1148
QY 390 -----TTLV-YKHLPAAPSAAPLPGIDSSALHPULCVGPEGQONLAPGACGCGDGT 441
D 1149 LEVEERTTILSGPRSKRASPILSPEP--QTPSLAPAKCSAPSSAVTPVPPA--SLIGSGT 1204
QY 442 DVLKCTHCAAAFHMRCHFPACTSRPGTGLRCRSCSGDVTAPAV--EGVLAPSPARLAPGP 499
D 1205 -----KPOVGLTSCRPAEAVPPAGLTKKGVSSPAP----PGP 1237
```

RESULT 3
US-09-764-864-1254

```
; Sequence 1254, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764, 864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1254
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (21)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (290)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (464)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1254
```

Query Match 5.7%; Score 166; DB 10; Length 549;
Best Local Similarity 24.9%; Pred. No. 0.0062;

Matches 61; Conservative 31; Mismatches 93; Indels 60; Gaps 10;

```
QY 115 PAVPALVP-----PRLPTKKAASEARAAPALTPRGTAAPGSLAKAPPKKRESSAQR 174
D 334 PPGPTSTANPSP-----GPSALSP--GSSGLNSHTPPVRRPSTSTSGSR 375
QY 175 LPLNGIQTMSAVQRAVAMSSGDV--PGARGAV--EGILIQVFESSGSKCQIY 226
D 376 GSCGSSGRT---AEKTSLSFKSDQYKVKQEPGTEDICFSFGVCKQEKEDGRSACMLS 432
QY 227 GGE-----FTTPSKFEDSGGKMKAR-----SSSGPKPLVRAKAQAAP 266
D 433 SPESLTPPLSTNLLESELALASLENNVKKXPADMNESCQGLSLVNGKSPIRSLM 492
QY 267 GGEARLGGQGSVPAPLALPSPDQLHQKNEDECAVCRDGGELICDGCPRAPHLACTSP 326
D 493 -HRSARIGDGN-----NKDDP-----NEDWCACVQNGGDLCCCEKCPKVFHTCHVPT 541
QY 327 LREIP 331
D 542 LLSFP 546
```

RESULT 4

```
US-10-005-057A-31
; Sequence 31, Application US/10005057A
; Patent No. US20020170087A1
; GENERAL INFORMATION:
; APPLICANT: Tao, Yunlin
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Shen, Bo
; APPLICANT: Lowe, Keith S.
; APPLICANT: Danilevskaya, Olga
; APPLICANT: Mahajan, Premod
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Sakai, Hajime
; APPLICANT: Klein, Ted M.
; TITLE OF INVENTION: Transcriptional Regulator Nucleic Acids,
; FILE REFERENCE: Polypeptides and Methods of Use Thereof
; CURRENT APPLICATION NUMBER: US/10/005, 057A
```

; CURRENT FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 60/251,555
 ; PRIOR FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 31
 ; LENGTH: 167
 ; TYPE: PRF
 ; ORGANISM: Oryza sativa
 ; US-10-005-057A-31

Query Match 5.7%; Score 165; DB 9; Length 167;
 Best Local Similarity 55.8%; Pred. No. 0.002;
 Matches 24; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 298 ECACVRDGEIICDGCRAFHACISPLREIPSGTRCSCC 340
 DB 105 ECVECDLGNLCCDSCRTYHLECLNPLKRAPPGNMOCPCRC 147

RESULT 5

; US-09-893-519A-14
 ; Sequence 14, Application US/09893519A
 ; Publication No. US20030027243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ANADYS PHARMACEUTICALS, INC.
 ; APPLICANT: THOMPSON, Craig
 ; APPLICANT: MOORE, Jeffrey
 ; APPLICANT: BURMAN, Ed T.
 ; APPLICANT: BRADLEY, John
 ; APPLICANT: DESILVA, Thamara
 ; APPLICANT: HARRIS, Sandra
 ; APPLICANT: KOMARNITSKY, Svetlana
 ; APPLICANT: MENDILLO, Marc
 ; APPLICANT: MOORE, Daniel
 ; APPLICANT: SANDERSON, Karen
 ; APPLICANT: HAQ, Tariq
 ; APPLICANT: ZHU, Shuhao
 ; APPLICANT: LONG, Fan
 ; APPLICANT: DAVIDOV, Eugene
 ; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
 ; FILE REFERENCE: 0342/1G548-US2
 ; CURRENT APPLICATION NUMBER: US/09/893,519A
 ; CURRENT FILING DATE: 2001-06-28
 ; PRIOR APPLICATION NUMBER: US 60/215,164
 ; PRIOR FILING DATE: 2000-06-29
 ; PRIOR APPLICATION NUMBER: US 60/224,457
 ; PRIOR FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 146
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 14
 ; LENGTH: 1023
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
 ; DATABASE ENTRY DATE: 1997-06-25
 ; RELEVANT RESIDUES: (1)..(1023)
 ; US-09-893-519A-14

Query Match 5.5%; Score 159.5; DB 9; Length 1023;
 Best Local Similarity 23.2%; Pred. No. 0.03;
 Matches 109; Conservative 26; Mismatches 163; Indels 171; Gaps 22;

QY 126 PRLTTRKRASEARAAPAL-----TPRGT--ASPSQAKAPPKPPKPESSAEGORUP 176
 DB 45 PRP-----EVRAAAGALGNHVVSSPAGACAGAPRAEPGAPGAAPPEPPACRAP 97
 QY 177 LNCIGTWSASVQRAVAMSGDVPCA---RGAVEGILIOQVRESGG-----SKKCIQVG 227

DB 98 GGGGQRRGPPPPRRPLVPAGPAPPAKLRPPPEG-----SAGACAPVPAAYAAVAG 149
 QY 228 GEFTYPSKFEDSGSGKKNKARSSGPKPLVRAKGAAGAPGGGEALGGQGSV----- 279
 DB 150 PEPAAPGAPKAGPAAALAAAGPGGPGP-----GPGPGPGPKPAGPAGATQINGSA 201
 QY 280 -----PA---PLALPSDPOLHOKNEDECAVCRDGGELICCDPCPRAF 318
 DB 202 LNSHHAAPAVSLVNNGPALLLPKPAAP-----GTVI-----QTRP 240
 QY 319 HLAISPLPRLREIPSGTRCSCCLOATVOEVPRAEPPRPQEPVETPLPGLRSAGEEVR 378
 DB 241 FVGAAPAPPAAPAPB-----PPAA--PAPAAAPAPPPPP--PAPATLA 279
 QY 379 GPEPEIAGMDTTLVYKHLPAAPSAAPLPGDSSALHLLCVPGGQNLAPGARGCVG 438
 DB 280 RPEGHP-AG-----PPTAAP-----AVPPAAANQGSAGAAPAPAPAGG 319
 QY 439 DGTDLRCHCAAFHWRCHPAGTS--RPGTGLRCRSCSGDVTAPAVEGVLAAPSPARL-- 495
 DB 320 -----PAGVSGPGPGA-----AAAPAGVRAESPKRVVQ 350
 QY 496 APGPAKDDTASHEPALHRDLSELTHTPDGILQWAIQSMARPAAPP 544
 DB 351 AAPPAOTPLAASGPA-----STAAASWVIGPTMGALPSPAAPP 392

RESULT 6

; US-09-764-864-1184
 ; Sequence 1184, Application US/09764864
 ; Patent No. US20020132753A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT223
 ; CURRENT APPLICATION NUMBER: US/09/764,864
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1792
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1184
 ; LENGTH: 432
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (31)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (35)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (38)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (40)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (44)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (102)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; US-09-764-864-1184

Query Match 5.4%; Score 156; DB 10; Length 432;
 Best Local Similarity 22.3%; Pred. No. 0.019;
 Matches 65; Conservative 39; Mismatches 100; Indels 88; Gaps 12;

QY 111 GKRRPAPVAPALVPPPLPTKRKASEARAAPALTPRGTAAGSOLKAPPKPPKPESSA 170
 DB 101 RKKKPDSPVPLPEPNKQEKTEKEERTNVGRTLRSPR-----ISRRTAVVAIEIDQKA 155

```

QY 171 EQORLPLGNG---IQTMSAVORA-----VAMSGDVGPARGAVEGILIQVFESE----- 218
Db 156 DKR---GEGEDEVEEESTALOKTKKELKSEKDTNSK-----VSKVPPKGVW 204
QY 219 -GSKKCIQVGEFYTPSKFEDSGSKNKARSSGKPLVRAGAOGAAGGEGARLGOOG 277
Db 205 TGSR---TRGRWKYSNDESESGSEKSSAASEEKESE-----EAILADDD 250
QY 278 SVAPRALPSPOLHOKNEDECAVCRDGLICDGCPRAFHACLSPLREIPSGTWRC 337
Db 251 EPCKKGLPNHPEL-----ILLCDSCSGYHTACTLRPLMIIPDGEWFC 294
QY 338 SSC-----LQATVOEV-----QPRAEPR-----POEP 360
Db 295 PRCOKHLCEKLEBQLODLVALKKKERAKERKERYVVGISIENTIIPPOEP 346

```

```

RESULT 7
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. US20020127553A1
; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
; APPLICANT: ROHN, Jennifer Leigh
; APPLICANT: WEISS, Bertram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 4725US
; CURRENT APPLICATION NUMBER: US/09/764, 176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the nuc
; OTHER INFORMATION: Ietc acid sequence of AAP-
US-09-764-176-7

```

```

Query Match 5.3%; Score 155; DB 10; Length 1400;
Best Local Similarity 22.4%; Pred. No. 0.078;
Matches 65; Conservative 38; Mismatches 99; Indels 88; Gaps 12;

```

```

QY 113 KRPRNPALVPPRLPTKKAASEBARAALTPRGASPGSOLAKPKPKPSSAEQ 172
Db 703 KRDPSPKVALEPENKQEKTEKEEKTNGRTLRSR-----ISRPTAKVAEINDQAKD 757
QY 173 ORLPLGNG---IQTMSAVORA-----VAMSGDVGPARGAVEGILIQVFESE-----G 219
Db 758 KR---GEGEDEVEEESTALOKTKKELKSEKDTNSK-----VSKVPPKGVWMTG 806
QY 220 SKKCIQVGEFYTPSKFEDSGSKNKARSSGKPLVRAGAOGAAGGEGARLGOOGSV 279
Db 807 SR---TRGRWKYSNDESESGSEKSSAASEEKESE-----EAILADDDP 852
QY 280 PAPRALPSPOLHOKNEDECAVCRDGLICDGCPRAFHACLSPLREIPSGTWRCSS 339
Db 853 CKKGLPNHPEL-----ILLCDSCSGYHTACTLRPLMIIPDGEWFCPP 896
QY 340 C-----LQATVOEV-----QPRAEPR-----POEP 360
Db 897 COKHLCEKLEBQLODLVALKKKERAKERKERYVVGISIENTIIPPOEP 946

```

```

RESULT 8
US-10-189-971-20
; Sequence 20, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:

```

```

; APPLICANT: walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kieelin-like Proteins and Polynuc
; FILE REFERENCE: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-20

```

```

Query Match 5.3%; Score 153; DB 9; Length 1207;
Best Local Similarity 22.7%; Pred. No. 0.088;
Matches 99; Conservative 44; Mismatches 164; Indels 130; Gaps 22;

```

```

QY 121 ALVPPRLPTKKAASEBARAALTPRGASPGSOLAKPKPKPSSAEQORLPLGNG 180
Db 7 SLCPARQCP---ADTQARSLGSAALSAVSTRTDTSIRARRPDSKSGASVSAAP----- 58
QY 181 IQTMSAVORAVVAMSGDVG-----GARGAV-----EGILIQVFESESKKCIQVYG 228
Db 59 -ARLARSVPKRSKAGSBPVPCPLLAASSAGACCPCSDCTYHSQVYANG----- 107
QY 229 EFTYPSKFEDESGSKNKARSSG-----PKPLVRAKGAOG--AAPGGEARLGOOG 277
Db 108 -----ONFTDADSPCHACHQODGTVCSLVDCPRTTCARQSGPCCRCPCDILEEYV 162
QY 278 SVAPRALPSPOLHOKNEDECAVCRDGLICDGC-----CPRAFHACLSPL----- 325
Db 163 EV-----DSEFSHPDRPCQECR-----COEGHANCOPRPCPRA---PCANPLRGT 205
QY 326 -----PLREIPSGT-----WRCSSCLQATVOEVQPR-----AEEP----- 355
Db 206 CCPNDCSGCAFGEKTPSADPRHPDPCRLCLSGNVQCLARRCVPLPCPREVLLPGE 265
QY 356 -RPQEP---PVETPLRPLRSAGEEVRGPRGEPPLAG---MDTTLVYKHLRPPS---AAP 405
Db 266 CCRQCPAARAPAGCPRPAGAAHNRHDEYFSPRQDRCRCCLDSGVSQRLPCRPACANP 325
QY 406 LRGDSSALHPLLYGPR---GQONLAPGARGC--VGGDGTDLVLRCTHCAAFNRCHRP 460
Db 326 RQGPCPCPCDGLYQGEKFASEGERPPYAAACHLICLWEGSVSCERKACAPAL---CPFP 382
QY 461 AGTSRPGTGLRCRCSG 477
Db 383 A-----RGDCCPDCDG 393

```

```

RESULT 9
US-09-819-104A-5
; Sequence 5, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
; FILE REFERENCE: UMG-030
; CURRENT APPLICATION NUMBER: US/09/819, 104A
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5

```

LENGTH: 2462
 TYPE: PRT
 ORGANISM: Mus musculus
 US-09-819-104A-5

Query Match 5.28; Score 151.5; DB 9; Length 2462;
 Best Local Similarity 22.98; Pred. No. 0.23;
 Matches 121; Conservative 42; Mismatches 186; Indels 179; Gaps 24;

98 DSFVKVDLSQPRKGRK-----PAYKALVPPRLPTKRASPEARAAAPALTPRGT 151
 746 ESVPSPSEAKKDTGPRKTGEALPAATQPPVPEEP-----AAAPAEPSVPD 795
 152 AS-PGSQLAKAPPKPPRESSA--EQORLPLNGIQTMSASVORAVA-----MSG 197
 796 ASGPSPSPSPAPAPATVVDQEAAPAPQTEDEAKQSEAEIDVCKPEEPSE 855
 198 DVP-----GARG--AVEGILIQV-FESGSKKCIQVGEFTYPSKFE 237
 856 EPPSPVSKDHKEETEEDKAKGTETVSEAPLKVEASKAAYTKGS--SSCATQ 912
 238 DSGS-----GKNKAR-----SSGPKPLVRAKAGGAA-- 265
 913 DSDSSATCSADEVDPEEGGKRLSPRSLITPAGDPRASTSPQRLDKOLKORAAAI 972
 266 -----PGGEARLQOGSVAPARLALPSDPLHOKNEDECAVCRDGEILICDGP 315
 973 PPIVTKVHEPP-----REDTVP-PKRVPPVPPPTQHLOREGVDSQSG-----GSP 1017
 316 RAFLIACISPLRLPSTGWRKSSCLQATVOEVOPRAEBRPOEPPVET----- 364
 1018 RGRSRSPVPAEKAEKAPFAPFAPTEGOSYRLSPHAGHRLSPHREVIKSTRADPLF 1077
 365 -----PLPRLGSAGEVEVGRPEGLAGMDTLVYKHLRPAAPRLPGDSSALHPL 417
 1078 SYTPPGHPLPLGLHDSANPV-----LPPRLPSNP-PLLISSAKHP- 1116
 418 LCVPEGOQLNLPARGCVCGDGTDV-LRCHCAAAFMHRCHEFPAGTSRPTGLRCRSCS 476
 1117 ---GVLEHQ-----LGAISQMSVQLRVPHSEHA-----KPMGPLMELPLA----- 1155
 477 GDVPAVEGVGLAPSPARLAPGAKDPTASHPALHDDLESLSERT 524
 1156 ---VDPKRLGTALAPPVEASP-----RASQYPGCRFPQLKRLYHPR 1195

RESULT 10
 US-09-925-301-844
 Sequence 844, Application US/09925301
 Patent No. US20020052308A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 FILE REFERENCE: PA106
 CURRENT APPLICATION NUMBER: US/09/925,301
 CURRENT FILING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: PCT/US00/05882
 PRIOR FILING DATE: 2000-03-08
 PRIOR APPLICATION NUMBER: 60/124,270
 PRIOR FILING DATE: 1999-03-12
 NUMBER OF SEQ ID NOS: 1694
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 844
 LENGTH: 601
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (36)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (64)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
 LOCATION: (103)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (106)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (152)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (358)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (383)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-301-844

Query Match 5.18; Score 149; DB 10; Length 601;
 Best Local Similarity 24.18; Pred. No. 0.073;
 Matches 121; Conservative 33; Mismatches 175; Indels 174; Gaps 27;

92 RLQPIIDSPFKVDN--LSQ-----PRKGRKPPAVKALVPPRLPTKR-- 133
 82 RLQVVIHSNPAQIQOALALQSLXROXSVTAPGHNHRKGPQAVQG--PSPRPRTYEPQ 139
 134 -----ASEEARAAAPAL-----TPRGTA--SP-GSQLAKPKPKPPSSAEQRLPIGN 179
 140 RVNSGLSSDHFKEPFGVMVGVGTPRDSAGVSFPFRKRRRPRKRLDLESILP--- 196
 180 GIOTMSASVORAVAMSSGDVPGARGAVEGILIQVFESGSKKCIQVGEFTYPSKFE 236
 197 -----PHSSGFL-GSK-PEG-----PGQAESNDTGTETALPHIMNRL 232
 237 EDGSGGKNKARSSGP-----KPLVRAKAGGAAPGGEALHGOQSVAPARLALPSDQL 291
 233 HTATSRKSYRSPMEPEMLSPEDVAGTEMSSDGSVDLSGSQVSSGCSQSSP-- 290
 292 HOKNEDECAVCRDGEILICDGCPRAFHCLASPLRLPSTGWRKSSCLQATVOEVOPR 351
 291 -----DGLKGAAG-----PPKR--PG-----SSLNAVCEGEPG 321
 352 AEPPRPOEP-----PVETPLPG-----LRSAGEEYGP-----PGEP-- 384
 322 SEPPRRRPPAPHDODRKELPREGLPPGPIGTERTSQXTDRGTE--GPIRPSHRCPPVQF 380
 385 -----LAGMDTLVYKHLPA-----PSAAPRLGLDSSALHPLCY 420
 381 GYXDKDSDLRLVVGDSLAKAEKELTASVTEALPVSRDWELPSAASAPQSKNLDSCGHV 440
 421 GPE---GQNLAPGARGCVCGDGTDVLRCTHCAAAFMHRCHEFPAGTSRPTGLRCRSCS 477
 441 -PEPSSSGQLRYPEVFGSAGPSSQI-----SGGAMSQQLHPNSGGFRPGT----- 486
 478 DVTAPAVEGVGLAPSPARLAPGPA 500
 487 -----PSLHPRYSQPLVLPGPBA 504

RESULT 11
 US-09-764-176-3
 Sequence 3, Application US/09764176
 Patent No. US20020127553A1
 GENERAL INFORMATION:
 APPLICANT: NOTEBOEN, Mathieu Hubertus Maria
 APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
 APPLICANT: ROHN, Jennifer Leigh
 APPLICANT: WEISS, Bertram
 TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
 FILE REFERENCE: 4725US
 CURRENT APPLICATION NUMBER: US/09/764,176
 CURRENT FILING DATE: 2001-01-17
 NUMBER OF SEQ ID NOS: 18
 SOFTWARE: PatentIn version 3.0

```

? SEQ ID NO 3
? LENGTH: 386
? TYPE: prt
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Description of Sequence: amino acid sequence of the analyzed regt
? OTHER INFORMATION: on of the Apoplin-associating clone AAP-2-II
? NAME/KEY: misc_feature
? LOCATION: (1)..(3)
? OTHER INFORMATION: Description of Sequence: the three C-terminal amino acids (His-Gln
? OTHER INFORMATION: u-Gly) are from the multiple cloning site of pACT
US-09-764-176-3

```



```

Db 66 PERTRAELEKLIQ--RAVL---RVSYSISYRNAARYQP-----PRRGAT 107
Qy 115 PPAVPA-----LVPPRLPTKKASEEA-RAAPALTP--RGTASP 154
Db 108 PPAAPRAARGAPAAAAAAPPPTTAPPPAPVAAAAAPARAAAAAATAPSPGAPQ 167
Qy 155 GSQK-----AKPKKPSSAEOQR-- 174
Db 168 GPRARAPALAAPPPAPAPAVAPAGRRAPPAVAREPPLPPPPAPQOQOP 227
Qy 175 -----LPLNGIQTMSAYORAVAM-----SSGDVPG--ARGAVEGILIQVEESG 218
Db 228 PPQPQPPPEGGAVAGAGAA--RPVSLREVRYLGGSGGAGRLTRGVGLLEEEAARG 285
Qy 219 GSKKCIQVGEFTYTPSKFEDSGSKNKARSSGPKPLVRAKGAOGAPG--GGEARLG 274
Db 286 -----RLKTRILG--ALALPRGDRP--GRAPPAASARPSKRGGEERYL 326
Qy 275 QGGSVPAPALPSPDQLHKNEDCAVCRDGGELICDGGPRAFHLACLSPLEIPIST 334
Db 327 EK-----EEEDDEDEDEDDVSGSEV-----PESDRPAG- 358
Qy 335 WRSSCLOATVQVOPRAAE--PRQEPVETPLPGLRSAGEEVRGP--PGE-----PLA 386
Db 359 -----AQHQLNGERGPQSAKERVKEMTPCGPHOGODEGRGAPAGSGTRQVFSMA 408
Qy 387 GMD-----TTLVY-----KHLPPAPSAAPLPGLDSSALHPLLVCVPGGOQ 426
Db 409 AMNKEGTAATGTPDPSPPVPLPGKPLPGADGT---PFGC--PPGRK 453

RESULT 14
US-10-023-523-43
; Sequence 43, Application US/10023523
; Patent No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-43

Query Match 5.18; Score 147.5; DB 12; Length 538;
Best Local Similarity 22.6%; Pred. No. 0.08;
Matches 106; Conservative 47; Mismatches 136; Indels 181; Gaps 26;
Qy 56 PQAETHALLSWLTDSTAILDFWVLFK-DYNLEERYGLQPLIDSPKDVLSQPRKGRK 114
Db 66 PERTRAELEKLIQ--RAVL---RVSYSISYRNAARYQP-----PRRGAT 107
Qy 115 PPAVPA-----LVPPRLPTKKASEEA-RAAPALTP--RGTASP 154

```

```

Db 108 PPAAPRAARGAPAAAAAAPPPTTAPPPAPVAAAAAPARAAAAAATAPSPGAPQ 167
Qy 155 GSQK-----AKPKKPSSAEOQR-- 174
Db 168 GPRARAPALAAPPPAPAPAVAPAGRRAPPAVAREPPLPPPPAPQOQOP 227
Qy 175 -----LPLNGIQTMSAYORAVAM-----SSGDVPG--ARGAVEGILIQVEESG 218
Db 228 PPQPQPPPEGGAVAGAGAA--RPVSLREVRYLGGSGGAGRLTRGVGLLEEEAARG 285
Qy 219 GSKKCIQVGEFTYTPSKFEDSGSKNKARSSGPKPLVRAKGAOGAPG--GGEARLG 274
Db 286 -----RLKTRILG--ALALPRGDRP--GRAPPAASARPSKRGGEERYL 326
Qy 275 QGGSVPAPALPSPDQLHKNEDCAVCRDGGELICDGGPRAFHLACLSPLEIPIST 334
Db 327 EK-----EEEDDEDEDEDDVSGSEV-----PESDRPAG- 358
Qy 335 WRSSCLOATVQVOPRAAE--PRQEPVETPLPGLRSAGEEVRGP--PGE-----PLA 386
Db 359 -----AQHQLNGERGPQSAKERVKEMTPCGPHOGODEGRGAPAGSGTRQVFSMA 408
Qy 387 GMD-----TTLVY-----KHLPPAPSAAPLPGLDSSALHPLLVCVPGGOQ 426
Db 409 AMNKEGTAATGTPDPSPPVPLPGKPLPGADGT---PFGC--PPGRK 453

RESULT 15
US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: Same
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-16

Query Match 5.0%; Score 145.5; DB 9; Length 1251;
Best Local Similarity 21.7%; Pred. No. 0.26;
Matches 97; Conservative 44; Mismatches 166; Indels 139; Gaps 21;
Qy 121 ALVPPRLPTKKRASEEAPRAAPALTPRGTAAPSGOLKAPPKPPSSAQOQLPLNG 180
Db 7 SLCPARQCP-----ADQASLSGSAALSAMAVSTDTITRARRPDSKASVSANP----- 58
Qy 181 IOTMSAYORAVAMSSGDVPGARGAVEGILI-----QQVEESG 219
Db 59 -ARLARSPVRSRQAQSPPYVCLPLAASAOQLTILSLPHGACRPSDCSTHSQVYANG- 116
Qy 220 SKKCIQVGEFTYTPSKFEDSGSKNKARSSG-----PKPLVRAKGAOG--AAAGG 268
Db 117 -----QNETDADSPCHACHCQDGTVCSLVDCPPTTCARPOSGPQCCPRC 162
Qy 269 GEARLGOQSVAPALPSPDQLHKNEDCAVCRDGGELICDGG-----CPRAFL 320
Db 163 PDCILEEYFV-----DGESFHPDRPCQECR-----CQGHACHQPRPCPRA--- 205

```

```

QY 321 ACLSP-----PLRETPSGT-----WRCSSCLQATVQEVQPR-----A 352
Db 206 PCAHPLPGTCCPNDCSCGAFGKKEYPFGADPHPSDPCRLCRLSGNVQCLARRCVPLPC 265
QY 353 EEP-----RQEP-----PVETPLPPGLRSAGEEVRGPPGEPLAG---MDTTLVYKHLP 398
Db 266 PEPVLLPGECCFQCPAPAPAPACCPRGAAHARHQEYFSPGDCRRLCLDGSVSCQRLP 325
QY 399 APPS--AAPPLGLDSSALHPLLCVGP---GOONLAPGARCG--VCGDSTDVLRCTHCAA 451
Db 326 CPAPCAHPRQGPCCPSCDGLYOGKEFASGERFPSPTAACHILCLCWEGSVSCQPKACAP 385
QY 452 AFHWRCHFPAGTSRPGTGLRCRSCSG 477
Db 386 AL--CPFPA-----RGDCCPDG 402

```

Search completed: March 13, 2003, 17:58:32
 Job time : 23 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:52:12 ; Search time 16 seconds
(without alignments)
1002.218 Million cell updates/sec

Title: US-09-509-595B-2
Perfect score: 2902
Sequence: 1 MARDALRLRLRLHREIAV.....DGLIQLMAIQSMARPAAPPS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	8.3	1912	4 US-08-913-832A-2	Sequence 2, Appl
2	241	8.3	1912	4 US-09-249-181A-2	Sequence 2, Appl
3	236.5	8.1	351	5 PCT-US91-06418-1	Sequence 1, Appl
4	213.5	7.4	753	4 US-08-942-686-2	Sequence 2, Appl
5	158	5.4	1593	4 US-08-628-829-4	Sequence 4, Appl
6	154	5.3	2972	4 US-09-579-181-2	Sequence 1, Appl
7	154	5.3	3118	4 US-09-579-181-1	Sequence 2, Appl
8	151.5	5.2	2289	3 US-09-051-019-2	Sequence 1, Appl
9	149	5.1	1384	4 US-08-976-255-11	Sequence 11, Appl
10	147	5.1	1093	3 US-08-545-860D-55	Sequence 55, Appl
11	147	5.1	1093	5 PCT-US94-04496-55	Sequence 55, Appl
12	144.5	5.0	902	1 US-08-396-479B-6	Sequence 6, Appl
13	144.5	5.0	902	1 US-08-818-823-6	Sequence 6, Appl
14	143	4.9	527	4 US-09-370-838-216	Sequence 216, App
15	142.5	4.9	510	4 US-08-246-489-2	Sequence 2, Appl
16	140.5	4.8	1274	4 US-09-095-443-2	Sequence 2, Appl
17	140	4.8	830	4 US-09-562-737-38	Sequence 38, Appl
18	139	4.8	1187	3 US-08-320-559-28	Sequence 28, Appl
19	139	4.8	1187	3 US-08-545-860D-28	Sequence 28, Appl
20	139	4.8	1187	5 PCT-US94-04496-28	Sequence 28, Appl
21	139	4.8	1210	1 US-08-320-559-26	Sequence 26, Appl
22	139	4.8	1210	3 US-08-345-860D-26	Sequence 26, Appl
23	139	4.8	1210	5 PCT-US94-04496-26	Sequence 26, Appl
24	138.5	4.8	1065	1 US-08-642-255-72	Sequence 72, Appl
25	137.5	4.7	1290	4 US-09-150-460B-6	Sequence 6, Appl
26	137	4.7	633	1 US-08-642-255-73	Sequence 73, Appl
27	136.5	4.7	830	4 US-09-562-737-33	Sequence 33, Appl

28	136.5	4.7	862	4 US-09-413-814-78	Sequence 78, Appl
29	135.5	4.7	335	2 US-08-405-175A-6	Sequence 6, Appl
30	135.5	4.7	2441	1 US-08-194-468-2	Sequence 2, Appl
31	135.5	4.7	2441	3 US-08-961-739-2	Sequence 2, Appl
32	135.5	4.7	2441	4 US-09-514-247A-8	Sequence 8, Appl
33	134	4.6	659	4 US-09-562-737-14	Sequence 14, Appl
34	134	4.6	992	1 US-08-127-499A-1	Sequence 1, Appl
35	134	4.6	992	1 US-08-482-847-1	Sequence 1, Appl
36	134	4.6	1185	4 US-09-041-886-23	Sequence 23, Appl
37	133	4.6	1317	3 US-09-083-521-7	Sequence 7, Appl
38	132	4.5	1004	4 US-08-916-352-2	Sequence 2, Appl
39	132	4.5	1050	4 US-09-428-711A-16	Sequence 16, Appl
40	130.5	4.5	1060	3 US-08-931-820-3	Sequence 3, Appl
41	130.5	4.5	1418	3 US-08-963-825-20	Sequence 20, Appl
42	130.5	4.5	1418	4 US-09-010-999-1	Sequence 1, Appl
43	130.5	4.5	1418	4 US-09-500-811-20	Sequence 20, Appl
44	130.5	4.5	1418	4 US-09-570-573-20	Sequence 20, Appl
45	130.5	4.5	1418	4 US-09-548-608-20	Sequence 20, Appl

ALIGNMENTS

```

RESULT 1
US-08-913-832A-2
; Sequence 2, Application US/08913832A
; Patent No. 6329517
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0030-999
; CURRENT APPLICATION NUMBER: US/08/913,832A
; CURRENT FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1912
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-913-832A-2

Query Match      8.3%  Score 241; DB 4; Length 1912;
Best Local Similarity 23.1%; Pred. No. 4e-11;
Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14;

QY 129 PTKRKASEEARAAAPALTPRGTAAPSGQLAKAPKPESSAEOQLPLGNGIQMTASAV 188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 PFKSGSASVAAAAMAAVAVESWVTADEV-APPPVEVPIRKAKTEGKGPNA----- 270

QY 189 QRAVAMSSGVDPKARGAVEGLILIOVFESG-----SKCIVGGFFTPSPFED----- 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 -RRPKSPRPDPADKKRPPKVPADLTKLGGFGSKRKSSSEDDLDVESPDFDASINSY 329

QY 239 SSGSKNKAARSSSGKPPVRAKGAAGAAAGGEGARLGOOGSPAPALPSPDLHOKNDE 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 SVSDGSTRSRSSRKRLRTK-----KKKGEE-EYTAUDGYETD-----HDDY 372

QY 299 CAVCRDGGELICDCGCPRAFLACLSPPLREIPSTGWRSCSLQATVOEVPRAEPPRO 358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 CEVCOOGEIILCDPTCRAYHMVCLDPMERKAPKWKSPHCEK---EGIOWEAKED--- 426

QY 359 EPPVEFLPGLRAGGEVRGPPGEPLAGMDTTLVYKHLPPAPSAAPLPGIDSSSLHLUL 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 -----NSGEELIEVVGDLLEEDD-----HHMEF----- 451

QY 419 CVPREGQNLAPGARCCVCGDGTDLNCTHCAAFHMRCHFPAGTSRGTGLRCSCSD 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 -----CAVCKKGGLLCCDITCPSSYHICLMPPLPEIPNGEMLCRCRCIP 496

QY 479 VTPAPVEGVL-----APSPARL-APGPAKDDTASHEPALHRDLESLSLSEHTFDGILO 531

```

Db 497 ALKGVOKILIMKQGPSPPTVPVRPPDADPNTSPKPLEGRERORFVK---WQGNSTW 553

RESULT 2

US-09-249-181A-2
Sequence 2, Application US/09249181A
Patent No. 6440679
GENERAL INFORMATION:
APPLICANT: Seelig, Hans Peter
TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
FILE REFERENCE: 8484-0059-999
CURRENT APPLICATION NUMBER: US/09/249,181A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 08/913,832
PRIOR FILING DATE: 1998-01-12
PRIOR APPLICATION NUMBER: PCT/DE96/00444
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1912
TYPE: PRT
ORGANISM: Homo sapiens
US-09-249-181A-2

Query Match

Best Local Similarity 8.3%; Score 241; DB 4; Length 1912;
Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14;

QY 129 PTRKASEEAAAPALTPRGTAAPGOLAKPKPKPSSAEOQLPLGNGIQTMASV 188
Db 217 PFKSSGSAVAAAAAAYVESMTATEV-APPPEVEYIRKKTDEGKPN- 270
QY 189 QRAVAMSSGDVPGARGAVEGLIQVEESG---SKKCIQVGEFFYPSKFE- 238
Db 271 -RRPKGSPVPPDKKPKPKKVAAPLKITLGFSGSKRRKSSSEDDLDVESPDASINSY 329
QY 239 SGSGKNKARSSSGPKPLVRAKGAAGAPGGEARLGOOGSPAPLALPSDQILKND 298
Db 330 SVSDGSTSRSSRSRKRTK-----KKKKEE-EYTAVDGYTD-----HODY 372
QY 299 CAVCRGGELICDCCPRAHILACSLPRLPISGTRWRCSSLOATVOEVPRAEPPRQ 358
Db 373 CEVQGGEGELLCDTTPRAYHMYCLDPMKAPGKWCSPCKE---EGIQWEAKED--- 426
QY 359 EPPVEPLPLGRLSAGEEVRGPPGELAGMDTLVYKHLPAAPSAAPLGLDSSALHPL 418
Db 427 -----NSEGEELIEVGGDLEEDD---HHMEF----- 451
QY 419 CVGPESQMLAPGARGCGDGTDLVRCYHCAAFHWRCHFPAGTSNRGTGLRCRSGD 478
Db 452 -----CRVCKDGGELLCDTCTPSYHICLNPLPLPELPCNEMWLCPRCTCP 496
QY 479 VTPARVEGL-----ASPARL-APGPAKDDTASHEPALHRDLSLSEHTFDGILW 531
Db 497 ALKGVOKILIMKQGPSPPTVPVRPPDADPNTSPKPLEGRERORFVK---WQGNSTW 553

RESULT 3

PCT-US91-06418-1
Sequence 1, Application PC/TUS9106418
GENERAL INFORMATION:
APPLICANT: Oklahoma Medical Research, Foundation, et al
TITLE OF INVENTION: Antigens Associated with Polymyositis
TITLE OF INVENTION: and with Dermatomyositis
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia

COUNTRY: US
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/06418

FILING DATE: 19910905

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/579023

FILING DATE: 09-JUL-1990

ATTORNEY/AGENT INFORMATION:

NAME: Padst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: OMRI120

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-572-6508

TELEFAX: 404-572-6555

SEQUENCE CHARACTERISTICS:

LENGTH: 351 amino acids

TYPE: AMINO ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

ORGANISM: Homo sapien

TISSUE TYPE: Sera

IMMEDIATE SOURCE:

CLONE: L1

PCT-US91-06418-1

Query Match

Best Local Similarity 8.1%; Score 236.5; DB 5; Length 351;
Matches 84; Conservative 25; Mismatches 131; Indels 85; Gaps 11;

QY 218 GSKKCIQVGEFFYPSKFE---DSGSGKNKARSSSGPKPLVRAKGAAGAPGGEAR-L 273
Db 8 GSKRRKGGSSDGPPEABESDLSGVSASGRPDGP---VTKTKLRGRPPKRRKKV 64
QY 274 GQSGSVAPALPSPOLKQKNEDECAYCRDGGELICDCCPRAHILACSLPRLPISG 333
Db 65 G-----CPVAGEEVDGYETDHDYCEVCGGGEITLCTCPRAHYLVCLDPELDRAPEG 120
QY 334 TWRCSSLOATVOEVPRAEPPRQEPPEVETPLPGLRSAGEEVRGPPGELAGMDTLV 393
Db 121 KWSCHKEKEGYWEAKKEEYEDE-----GEE-BEKEEE---DHME 161
QY 394 YKHLPAAPSAAPLPELDSALHPLLCVGPESQMLAPGARGCGDGTDLVRCYHCAAF 453
Db 162 Y-----CRVCKDGGELLCDTCTPSYHICLNPLPLPELPCNEMWLCPRCTCP 496
QY 454 HWRCHFPAGTSRPGTGLRCRSGDVTTPARVEGL-----ASPARLAPGPAKD 503
Db 183 HIHLNPLPLPPLNGEWLCPKCTCPVLKGRVOKILIMKQGPSPPTVPVRPPDADPNTSPK 531
QY 504 TASHEPALHRDLSLSEHTFDGILW 531
Db 240 VPPRPLQGRSEREFVK---WGLSTW 264

RESULT 4

US-08-942-686-2
Sequence 2, Application US/08942686
Patent No. 6183988
GENERAL INFORMATION:
APPLICANT: Bloch, Donald B.

Db 335 -----SAP-----APKGRSPSPSS----- 351

Qy 451 AAFHMRCPAGTS-RP-GTGRCSGSGVTPAPVEGVLAAPSPAPLAPGPAPADDTASHE 508
 Db 352 -----PSGRSVKPESEPGVRRKR-----VSPVPOSGRI-TPPRAPSP--DGFSPPS 395

Qy 509 P 509

Db 396 P 396

RESULT 6
 US-09-579-181-2
 ; Sequence 2, Application US/09579181
 ; Patent No. 6365372
 ; GENERAL INFORMATION:
 ; APPLICANT: Chiviva, John
 ; APPLICANT: Yaciuk, Peter
 ; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
 ; FILE REFERENCE: 16153-4247
 ; CURRENT APPLICATION NUMBER: US/09/579,181
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: 60/136,620
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 2972
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-579-181-2

Query Match 5.3%; Score 154; DB 4; Length 2972;
 Best Local Similarity 22.0%; Pred. No. 0.001;
 Matches 114; Conservative 44; Mismatches 195; Indels 166; Gaps 24;

Qy 102 KDVDLSQPRKGRKPPAV-----PK-ALVP-PPRLPTKRKASEARAAAPAA 145
 Db 1027 RQLAVGQPRQLQMPPTVMNNTGVYKIVRQAPRDGLTPVPLAPAPRPPS-----SGLPAAV 1082

Qy 146 LTPRGTSAPGSQAKAPPKPKPESSAEQRLPLNGIOTMSASVORAVAMSSGDV-----PG 201
 Db 1083 LNRPLTLTPG-----RLPPTLTGTARAPMPT--PTLVRLPLKLVHSPPEVSASAPG 1132

Qy 202 ARGAVEGILIQVEFESGSKCIQVGEFTTPSKFEDSGSGKNKARSSGPKPLVRAKGA 261
 Db 1133 AAP-----LTSSPLHVPSSLPGPASPMPIMPSS--PLASPVSS 1170

Qy 262 QGAAPGGGARLGGQGSVPAPLAP-----SDQLHQKNEDECAVCRDGGELICDGCPR 316
 Db 1171 TVSVPLSSSLPISVPTTLAPASAPLTIPIISAP-----LTVSASGAL----- 1213

Qy 317 AFHLACLSPPLREI-----PSGTWRCSSCIQATVOEVOAPRAEPPPOEPVETP 365
 Db 1214 ---LTSVTPPLAPVVAAPGPPSLQPSGASPSASALTGLGATAPSLSSSQTGPHLLAP 1270

Qy 366 LP---PGLRSAGEEVRGPGEPLAGMDTLVYKHLAP-----P 401
 Db 1271 TSSHVPGLNSTVAPACSPVLVPAASALASPPSPAPNPAPAQASILAPASSASQALATPLAP 1330

Qy 402 SAAP---LPGIDSSALHPLILCVGPEGQONLAPGARCGVCGDGTDLRCHCAAAAFHMRG 457
 Db 1331 MAAPQATILAPSPAPLAPLPLVLA-----SPGAAPVLASSQTPV----- 1370

Qy 458 HFPAAGTSRPGTGLRCRSCSGDVTAPAVEGVLAAS-----PARL---AGPAKDDTASHE 508
 Db 1371 PVMAPSSSTPGTSL-----ASASVPAPPTP-VLAPSSQTGTLMLPAPVSPPLSPASTQTALAA 1425

Qy 509 PALHRDLESLSSEHTFDGIIQMAIOSM--ARPAAPPS 545
 Db 1426 PAL-----APTLCGSSPSSQTLSTGTCNPGCPFPPT 1454

RESULT 7
 US-09-579-181-1
 ; Sequence 1, Application US/09579181
 ; Patent No. 6365372
 ; GENERAL INFORMATION:
 ; APPLICANT: Chiviva, John
 ; APPLICANT: Yaciuk, Peter
 ; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
 ; FILE REFERENCE: 16153-4247
 ; CURRENT APPLICATION NUMBER: US/09/579,181
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: 60/136,620
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 3118
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-579-181-1

Query Match 5.3%; Score 154; DB 4; Length 3118;
 Best Local Similarity 22.0%; Pred. No. 0.0011;
 Matches 114; Conservative 44; Mismatches 195; Indels 166; Gaps 24;

Qy 102 KDVDLSQPRKGRKPPAV-----PK-ALVP-PPRLPTKRKASEARAAAPAA 145
 Db 1173 RQLAVGQPRQLQMPPTVMNNTGVYKIVRQAPRDGLTPVPLAPAPRPPS-----SGLPAAV 1228

Qy 146 LTPRGTSAPGSQAKAPPKPKPESSAEQRLPLNGIOTMSASVORAVAMSSGDV-----PG 201
 Db 1229 LNRPLTLTPG-----RLPPTLTGTARAPMPT--PTLVRLPLKLVHSPPEVSASAPG 1278

Qy 202 ARGAVEGILIQVEFESGSKCIQVGEFTTPSKFEDSGSGKNKARSSGPKPLVRAKGA 261
 Db 1279 AAP-----LTSSPLHVPSSLPGPASPMPIMPSS--PLASPVSS 1316

Qy 262 QGAAPGGGARLGGQGSVPAPLAP-----SDQLHQKNEDECAVCRDGGELICDGCPR 316
 Db 1317 TVSVPLSSSLPISVPTTLAPASAPLTIPIISAP-----LTVSASGAL----- 1359

Qy 317 AFHLACLSPPLREI-----PSGTWRCSSCIQATVOEVOAPRAEPPPOEPVETP 365
 Db 1360 ---LTSVTPPLAPVVAAPGPPSLQPSGASPSASALTGLGATAPSLSSSQTGPHLLAP 1416

Qy 366 LP---PGLRSAGEEVRGPGEPLAGMDTLVYKHLAP-----P 401
 Db 1417 TSSHVPGLNSTVAPACSPVLVPAASALASPPSPAPNPAPAQASILAPASSASQALATPLAP 1476

Qy 402 SAAP---LPGIDSSALHPLILCVGPEGQONLAPGARCGVCGDGTDLRCHCAAAAFHMRG 457
 Db 1477 MAAPQATILAPSPAPLAPLPLVLA-----SPGAAPVLASSQTPV----- 1516

Qy 458 HFPAAGTSRPGTGLRCRSCSGDVTAPAVEGVLAAS-----PARL---AGPAKDDTASHE 508
 Db 1517 PVMAPSSSTPGTSL-----ASASVPAPPTP-VLAPSSQTGTLMLPAPVSPPLSPASTQTALAA 1571

Qy 509 PALHRDLESLSSEHTFDGIIQMAIOSM--ARPAAPPS 545
 Db 1572 PAL-----APTLCGSSPSSQTLSTGTCNPGCPFPPT 1600

RESULT 8
 US-09-051-019-2
 ; Sequence 2, Application US/09051019
 ; Patent No. 6103229
 ; GENERAL INFORMATION:
 ; APPLICANT: KAHMANN, Regine and QUADBECK-SEEGER, Claudia
 ; TITLE OF INVENTION: Regulatory gene from *Ustilago maydis*
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Keil & Weinkeuf
 ; STREET: 1101 Connecticut Avenue

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
COMPUTER: IBM AT-compatible, Pentium processor
OPERATING SYSTEM: Windows 98
SOFTWARE: Wordperfect version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,019
FILING DATE: 31-MAR-1998
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2289 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-051-019-2

Query Match 5.2%; Score 151.5; DB 3; Length 2289;
Best Local Similarity 22.1%; Pred. No. 0.0012;
Matches 84; Conservative 41; Mismatches 118; Indels 137; Gaps 19;

QY 28 LHALADHDVPEDEKQETLHLKEKGCQPAFHALLSWLTQDSTALIDFWKDYNL 87
DB 285 LUNSL-ADARASQNYOEGL-----OKFHA----- 308
QY 88 ERYGRLO--PILDSFKVDLSQPR-----KGRKPPAVPKAL-----VPPPR 127
DB 309 -QGQKKRVSVPYID--RSVDLYQLKLVISLIGYDAVCAKRRKSDATRKIGYSDKEGQ 365
QY 128 LPTKKAS-----EARRAAPALTPRGTAAPSGSLAKAPPKPPKSSAEQRLPLGN 179
DB 366 LSTQYKAAVTRIIIPFEELAKAKESRPNG--SSVSPQLAQSALMGATASTDQE---N 420
QY 180 GIO--TMSAVQRAVAMSGDVGARGAVGILIQVFESGSKKCIQVGEFTYPSKFE 237
DB 421 GVKHPSMSGLD--AAPSGD-----AGHFRTKPEPFT 450
QY 238 DSGSKNKRAS-----SGPKPLVRAKGAOGAAPGGGEARLGGQGSVP--APL 283
DB 451 AAGAALANATPVLTPQSPSTVASTRSKAKRSATSTPASSSRSLDITSTPMPL 510
QY 284 A-----LPSPQL-----HQRNDE-CAVC--RDGELICDGCPRAFHLA 324
DB 511 ISRRKGVSPHLEADSYLRAQAGNQAEQMCIELRGEDGPNMLLCDECRGYHMYCLQ 570
QY 325 PPLREIPSGTWRCSSCLQAT 344
DB 571 PALTSPKQWFCPPCLVGT 590

RESULT 9
US-08-976-255-11
Sequence 11, Application US/08976255
Patent No. 6136581
GENERAL INFORMATION:
APPLICANT: Jono, Keith E.
APPLICANT: Plozman, Gregory
TITLE OF INVENTION: KINASE GENES AND USES
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage

COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976, 255
FILING DATE: No. 6136581ember 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,675
FILING DATE: No. 6136581ember 22, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 229/182
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1384 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-976-255-11

Query Match 5.1%; Score 149; DB 4; Length 1384;
Best Local Similarity 23.6%; Pred. No. 0.00097;
Matches 96; Conservative 42; Mismatches 141; Indels 128; Gaps 21;

QY 101 PKVDLSQPRKRPAPVAVPKALVPPRLPTRKASEEARA--APALTPRGTAAPSGSLQ 159
DB 790 PK--LATEAGTGTGPHLPVSPSPSGEAPLSEKASAPAPALPPSPPTATGEVS 846
QY 160 A-----KPKPPSSAEQRLPLGNGIOTMSA-----SVQRA 191
DB 847 AIRLASALNGSSSPSEVEAPSSSEDEDTAEATSGIPTDSSDGLQARRPDVPAFRSLQK 906
QY 192 V--AMSGDVGAR--GAVGILIQVFESGSKKCIQV--GEFTYPSKFE--EDSGS 241
DB 907 VGTIPDSIDTIDTSSADGCTEVFPSPATGSPGGOPRALDSYDTNENESPEFLKAEQ 966
QY 242 G-----KNKARSSSGKPLVRAK-----GAOGAAGGGEARLG 274
DB 967 GCEPQAFELASGECEGPEFRRLSTSLGMEKNRYRDSATVSDLEAEALASGPEKKG 1026
QY 275 QQGSVAPALPLPSDPQLHQRNDECAVCRG-----GELICDGCPRAFHLA 324
DB 1027 GDRAPGPELGLPS--TGQPSQVY--LRPGVSGEAGSGPGEVL----- 1066
QY 325 PPLREIPSGTWRCSSCLQATVQVQRAEP--RQGEPP-----VETPLPGLSAG- 374
DB 1067 PPLQLBEGSSPEPSTCPSGLVPE--PPEQGPAAKVPSPSCSQPFLLPV--LRSEGN 1123
QY 375 -EEVRGPPG-----EPLAGMDTLYYKHLPAAPPASAPPLGGLDS 412
DB 1124 SSEFGQPGLLSGPAPQKRMGCGPT-----PRAPLALPLGLRAA 1163

RESULT 10
US-08-545-860D-55
Sequence 55, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Regio
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &

```

Query Match      5.1%; Score 147; DB 3; Length 1093;
Best Local Similarity 18.7%; Pred. No. 0.001;
Matches 113; Conservative 77; Mismatches 249; Indels 166; Gaps 20;

QY      38 VPEKRFQETLHKKE-----KEGRQAFH---ALLSWLLTQDSTAIIDF 77
      || : : : : : : : : : : : : : : : : : : : : : : : :
Db      111 VPDRFNNTCYICETGREGSKAASGACMTCNHHGCRQAFHVTCAOMAGILCEE--VLEV 168
      || : : : : : : : : : : : : : : : : : : : : : : : :

QY      78 WRVLFKDYNLERYGLQDILDS-----PFRVDL 106
      | : : : : : : : : : : : : : : : : : : : : : : : :

Db      169 DNIVYCYGCKYHFSMKTSRRSHSGGGGGGAGGGGSGMGSGGSGFTSGRRSRASSTQDE 228
      | : : : : : : : : : : : : : : : : : : : : : : : :

QY      107 SQP-----RKGRKPRVAKVLPVPPRLPTKKRASEEARAAAPALTP 148
      | : : : : : : : : : : : : : : : : : : : : : : : :

Db      229 KHPHNEHGQKSRKDKERLQKHKKRPESPPSITLPVVLADKVVSSASSSSHHNEAST 288

```

```

QY      149  RGAASGSOAKAPPKPKFESSAEQORPLPIAGIOTMSASVQRAYVMSSGDVPGARGAVBG 208
Db      289  QETSESSREKGGKSSSHSLSHKCKKLSSCGVCSFTTSASSSSSSSSSGGPPQPAVSS 348
QY      209  I-----LIQVFEESGGSKKCIQVGEFFTPS-----K 235
Db      349  LQSPDFSAAPKLEQPEEDYKSKPTAPAPAPSPSAPEPPKADLEQKVVESSGFPIMR 408
QY      236  FEDSGSGKKMKARSSSG---PKPLVRAGAGAAAGCGGARLGGQGSVAPLALPDPOLH 292
Db      409  FSTTTSSSGRARAPSPEDYKSPHYTSGAS---AGTHKRPALSAFTVPADEPPEGTGLK 464
QY      293  QKNEDECAVCROGELICDCGCPAFH-----LACTPLPLREIPSGT-----WRCS 339
Db      465  EKKHKASKRSRHH-----PGRPKGSNNKEGTGCPAPSLPSAQLAGTATTAASFSGS 518
QY      340  CLOATYQVQPAPEERPOPE--PVETP-LPGLRSAGEEYRGPPGEPLAGMDTTLVYKH 396
Db      519  LVSSGGGJSSRTFFGSGSLPSLSTLESPLTAGAGIYTSNKPISHSGCCMLRAVCSSTLSS 578
QY      397  LPAPSAAPLPGDSLALHPLLCVGPQGQNLAPGARGCVGGDGTDLVIRCHCAAFHWR 456
Db      579  ILGPGTSLALPLRSR---PFTSLPSSSASISTYQVFLAG-STFSLPSHIFTG---- 630
QY      457  CHFPAQTSRP-----GTGLRCRSCSGDVTAPAPVGVYLAAPSPARLAPGPAK 501
Db      631  ---PQGAVPNPLSQAESHTPELDCEFSRCRGS-----PQSLSMSPTISSLPALF 680
QY      502  DDTAS 506
Db      681  DDTAS 685

RESULT 11
PCT-US94-04496-55
: Sequence 55, Application PC/TUS9404496
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESSEE: Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TDU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1093 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

```


MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 PCT-US94-04496-55

Query Match 5.1%; Score 147; DB 5; Length 1093;
 Best Local Similarity 18.7%; Pred. No. 0.001;
 Matches 113; Conservative 77; Mismatches 249; Indels 166; Gaps 20;

QY 38 VPEDEFOETLLAKE-----KEGCPQAFH---ALISWLLTODSTALIDF 77
 DB 111 VPHDFKTKTCICETBETGESKASACMTCNHCGRQAHVYCAQAGLICEE--VLEV 168
 QY 78 WRVLFKDYNLERYGRLOPILDS-----FPRKDVL 106
 DB 169 DNYVYCGYCKYHFSKMKTSRHSSSGGGAGGSGGSGFISGRRSASASTQOE 228
 QY 107 SQP-----RKGRKPRVAVKALVPPRLPTKRKASEEPAAPALTP 148
 DB 229 KHPHNERGQKSKRKDEKRLQKHKKRPESSPILTPPVPTADKYSSSSSSSHHEAST 288
 QY 149 RGTAPESSQLAKPKPKRESSAEQORLPLGNGIOTMSASVORAVAMSGDVPAGAVEG 208
 DB 289 QETSESSRESKCKSSSHSLSHKGLSSKGVSTTSASSSSSSSGGFPQPAVSS 348
 QY 209 I-----LIQVFESSGSKKCIQVGEFEYTPS-----K 235
 DB 349 LQSSPDESAFPEKLEQPEEDKYSKPTAPAPASAPSPSAPPEPKADLFQKVVFSGFPIR 408
 QY 236 FEDSGSKNKARSSSG---PKPLVPAKAGQAGAPGGGEARLQGGGVPAFLPSPQLH 292
 DB 409 FSTTSSSGRRAPSPGDKSPHVTGSGAS---AGTHKRMPLASTATPVADETPETGLK 464
 QY 293 QKNEDECAVCDGELLICDCCPRAFH-----LACLSPLEIREISGT-----WRCS 339
 DB 465 EKKKAKKRSRHS---PGRPKSRNKEGTGCGRAPSLPSQLAGFTTAASPEGGS 518
 QY 340 CLQATVEVOPRAEPRQEP--PVETP-LPPGLSAGEEVRGPPEPLAGMDTTLVYKH 396
 DB 519 LVSSGLGLSRTGCPGSLPSLSLEPLGAGIYTSNKPDISHSGMLRACVSTPLSS 578
 QY 397 LPAPSAAPLPGLDSSALHPILCVPEGOQNLAPARCGVCGDGTIVLCTHCAAFHWR 456
 DB 579 LAGPPTGALSRLRS---PETSTLPSASSASISTQVFSLAG-STPSLSTHIFGT--- 630
 QY 457 CHEPAGTSRP-----GTGLRCSGSDVTPAPVEGLAPAPALPAPAK 501
 DB 631 ---PMGAVNPLLSQAESHTPELDCEFRCKGTS-----FOESLSMSPISSLPA-LF 680
 QY 502 DDTAS 506
 DB 681 DDTAS 685

RESULT 12
 US-08-396-479B-6
 Sequence 6, Application US/08396479B
 Patent No. 5612455
 GENERAL INFORMATION:
 APPLICANT: HOEY, Timothy
 TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/396,479B
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-59450-1/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-8700
 TELEFAX: (415) 494-8771
 TELEX: 210 277299
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 902 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-396-479B-6

Query Match 5.0%; Score 144.5; DB 1; Length 902;
 Best Local Similarity 23.7%; Pred. No. 0.0013;
 Matches 113; Conservative 29; Mismatches 136; Indels 199; Gaps 25;

QY 113 RKPAPVPAKALVPPRLPTKRKASEEPAAPALTPPGTASPGSQ-----LKA 160
 DB 63 RPPRPGRHSPPR-PAPSPGTWESGPARSVRLGCGGAGGAGGVLECPSTRITSI 121
 QY 161 KPKKRESSAEQORLPLGNGIOTMSASVORAVAMSGDVPAGARAVEGILIQVFESSGS 220
 DB 122 SPTPEPALDENPDAMGDG-----SPRDYPPPEG-----FGGY 155
 QY 221 KKC-IQVGEFEYTPSK-----FEDSG-----SGKNKARSSSG---PK 253
 DB 156 KAGAGGAGGAFSPSGSSLSWSFFSDASDEALTYAACDEVESELEAASRGLGSP 215
 QY 254 PLVRAR-----GAQGAAPG-----GEAR----- 272
 DB 216 PEPRAAPRPWTEPDPSLVGSPGGRGREDSWLLSAPGTPPASPRASPGCKRYS 275
 QY 273 -----LQOGGS---VPAPLALPSPQLHOKNEDECAVCDGELLICDCGCP 315
 DB 276 TPSSASPALSRGSLCEEGSEPPPLPLARDP-----GSP 312
 QY 316 RAFLHACLSPLREITSGTWRCSSCLQATYQVOPRAEPRQEPVETPLPGLRSAGE 375
 DB 313 GPFIDYVG-APPAESTIPQKTRRIS-EOAV--ALPRSEEP---ASCNGKLPILG---AE 359
 QY 376 EVRGPPG--EPLAGMDTTLVYKHLPAAP--SAAPLPG---LDSALHPLLCVGEQO 426
 DB 360 ESVAPPGSRKEVAGMD-----YLAVPSPLAMSKARIGHSPIFTLSALPLDWPPLPSYE 415
 QY 427 NLAPGARGVCGDGTIVLCTHCAAFHWRCHPEPAGTSRPTGLRCSGSDVTPAP 483
 DB 416 QLE-----LRLEVOPRAHN-RAHYETEGSR-----GAVNAAP 446

RESULT 13
 US-08-818-823-6
 Sequence 6, Application US/08818823
 Patent No. 5708158
 GENERAL INFORMATION:
 APPLICANT: HOEY, Timothy
 TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818, 823
FILING DATE: 14-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/396, 479
FILING DATE: 02-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59450-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 902 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-818-823-6

Query Match 5.0%; Score 144.5; DB 1; Length 902;

Best Local Similarity 23.7%; Pred. No. 0.0013;

Matches 113; Conservative 29; Mismatches 136; Indels 199; Gaps 25;

QY 113 RKPAPVAKLVPPRLPTKRKASEARAAPALTPRCTASPGS-----LKA 160
DB 63 RPPPRGMSHPPR-PAPSPGTWESQPARSVRLGPGGAGAGGVLECPSTRITSI 121
QY 161 KPRKKPESSAEQRLPLGNGIQTMSASVQRAVAMSSGDVPGARGAVEGILIQVFESGGS 220
DB 122 SPPEPPALEDNPDWDG-----SPRDYPPPEG-----FCGY 155
QY 221 KKC-IQVGEFTYPSK-----FEDSG-----SGKKKASSSG---PK 253
DB 156 REAGAGCGGAFSPGSSSLSSMSFTSDASDEALYACDEVESELENEASRGGLGSP 215
QY 254 PLVRAK-----GAQGAAPG-----GEAR-----272
DB 216 PSPRASPRWTPEDPMSLVGSPGGRGPEDESWLLSAPGTPRASPRAFGKRRYSWG 275
QY 273 -----LGGQGS-----VPAPLALPSPDLHOKNDECAVCARDGELICDGC 315
DB 276 TSSSASPALSRRGSLGEEGSEPPPPPLPLARDP-----GSP 312
QY 316 RAFLACLSPPLREIPSGTWRCSCLOATVQVQRAEPRRQEPVETPLPGLBSAGE 375
DB 313 GFEDVYG-APRAESTIQKRRITS-EQAV---ALPREEP-----ASNGKLPUG---AE 359
QY 376 EYRGPRG---EPLAGMDTLVYKHLRAPP---SAAPLPG---LDSALHPLLCGPEGOQ 426
DB 360 ESVAPPGSKRKEVAGMD---YLAVSPPLAMSKARIGCHSPIRTGALLPLDMLPDSQYE 415
QY 427 NLAPGARGCVCGDGTVDVLCHTCAAFHMRCHFPACTSRPGTGLRCRSCGSDVTPAP 483
DB 416 QLE-----LRIEIVOPRAHH-RAHYETEGSR-----GAVKAAP 446

RESULT 14

US-09-370-838-216
Sequence 216, Application US/09370838
Patent No. 6444425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamatch, Roadonh

APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
FILE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370, 838
EARLIER APPLICATION NUMBER: 1999-08-09
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 216
LENGTH: 527
TYPE: PRT
ORGANISM: Homo sapien
US-09-370-838-216

Query Match 4.9%; Score 143; DB 4; Length 527;

Best Local Similarity 23.8%; Pred. No. 0.00085;

Matches 126; Conservative 36; Mismatches 191; Indels 176; Gaps 29;

QY 105 DISQPRKG-----RKPAPVAKLVPPRLPTKRKASEARAAPALTPRCTASPG 155
DB 30 ELTPPREKALENGELRSPDAGEKVLVNGILTPPK---SEDKYSENGLAEPRTTERP 86
QY 156 SOLKAKP-----KPESSAEQRLPLGNGIQTMSASVQRAVAMSS-----CD-VPGAR 203
DB 87 ETGPMWAPGEMKTPESWGAPPT---GEPAPETSLERAPAPSAVSSRRNGETAGP 142
QY 204 GAV-EGILIQVFESGSKKIQVGEFTYPSKE-DSSGKKKASSSGPKLVAKG 260
DB 143 GRAPKMGTL-----EPTERRAPETGAPAPAGRGLDLSG-----RAPV 184
QY 261 AGAAPGCG-----EARLGGQGSV---PAPLALPSPD-----POLHOKNE 296
DB 185 GTGTAPAGCGPGSGVDKAGAVDNTTRPQPPPLPPEEAPRRLRAPRAPEVAPEGE 244
QY 297 DECAVCRDGE-LICDGE-----CPRAFLACLSPPLREIPSGTWRCSCLOAT 344
DB 245 PGAPDSRAGDVALSGDGPPEPKERKPEMPRLF-LDGLPP-----QGN 287
QY 345 VOEVOPRAEPRRQEPVETPLP-PGLR-----SAGEVNGPPEPLAGMDTL 392
DB 288 SEQIKARLSRLALPPLTLTPPGCGPRRPWEGADAGAGGAGC-----AG-----336
QY 393 YKHLRPAPEAPLPGDLSALHPLLCVPGEQNLAPARCVCGDGT-----VLR 445
DB 337 -----APPAEEDGEDEDE-----EEDBEAAPGAAAGPRGRRARAPVYVS 383
QY 446 CTHCAAFHMR---CHFPAGTSRPGTG-----LRCNSCGDVT-----PAP 483
DB 384 SADADAAPRLKGLKSPGADPEDESELERKKRMVSHGDVYVYLPDQETPTNELSVQAP 443
QY 484 VEGVLAPSPARLAPGPAKDDTASHHEPALHRODLESLSHTDGLIOMA 532
DB 444 PEGDIDPESTPAPRTPP-----HPATPGDGPPS---NDSGFGSGFEWA 483

RESULT 15

US-08-246-489-2
Sequence 2, Application US/08246489
Patent No. 6225049
GENERAL INFORMATION:
APPLICANT: Lan, Michael S.
APPLICANT: No. 6225049kins, Abner L.
TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobe, Matens, Olson & Bear
STREET: 620 Newport Center Drive
CITY: Newport Beach
STATE: California
COUNTRY: USA
ZIP: 92660

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:50:47 ; Search time 14 Seconds

(without alignments)
1614.614 Million cell updates/sec

Title: US-09-509-595B-2

Perfect score: 2902
Sequence: 1 MAIDALRLRLRLHRTETAV.....DGLQWAIQSMARPAAPFPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2902	100.0	545	1	AIRE_HUMAN
2	2005	69.1	552	1	AIRE_MOUSE
3	270.5	9.3	1944	1	CHD3_HUMAN
4	241	8.3	1912	1	CHD4_HUMAN
5	228	7.9	689	1	SP11_HUMAN
6	227	7.8	1050	1	TEF1A_HUMAN
7	227	7.8	1982	1	CHDM_DROME
8	217	7.5	1051	1	TEF1A_MOUSE
9	206	7.1	1787	1	CHD3_CAERL
10	204.5	7.0	882	1	LY10_HUMAN
11	204	7.0	1127	1	TEF1G_HUMAN
12	186.5	6.4	607	1	YAZ7_SCHPO
13	179.5	6.2	404	1	YACS_SCHPO
14	168.5	5.8	660	1	YH11_EBV
15	161	5.5	684	1	YH11_EBV
16	160	5.5	835	1	YH11_EBV
17	159.5	5.5	1083	1	TEF1B_HUMAN
18	157	5.4	834	1	TEF1B_HUMAN
19	152.5	5.3	2161	1	TEF1B_MOUSE
20	152	5.2	366	1	SHK1_HUMAN
21	152	5.2	380	1	REO2_XENLA
22	152	5.2	810	1	REO2_CHICK
23	152	5.2	1049	1	REO2_CHICK
24	151.5	5.2	728	1	CA13_BOVIN
25	149.5	5.2	2167	1	YH89_YEAST
26	147.5	5.1	476	1	SHK1_RAT
27	147.5	5.1	2715	1	TRX2_HUMAN
28	147	5.1	1093	1	TRX2_HUMAN
29	146	5.0	2517	1	AF17_HUMAN
30	144.5	5.0	902	1	NCR4_HUMAN
31	144.5	5.0	1183	1	NCR4_HUMAN
32	144	5.0	388	1	DRPL_RAT
33	143.5	4.9	923	1	REO1_XENLA
					PRGR_MOUSE

ALIGNMENTS

RESULT 1
ID: AIRE_HUMAN
AC: 043918; 043932; 043922; 075745; PRT; 545 AA.
DT: 15-JUL-1998 (Rel. 36, Created)
DT: 15-JUL-1998 (Rel. 36, Last sequence update)
DT: 15-JUN-2002 (Rel. 41, Last annotation update)
DE: Autoimmune regulator (Autoimmune polyendocrinopathy candidiasis
ectodermal dystrophy protein) (APECED protein).
OS: Homo sapiens (Human).
OC: Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC: Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX: NCBI_TaxID=9606;
RN: [1]
RS: SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND VARIANT APECED GLU-83.
RC: TISSUE=Thymus;
RX: MEDLINE=98061086; PubMed=9398839;
RA: Nagamine K., Peterson P., Scott H.S., Kudoh J., Minoshima S.,
RA: Heino M., Krohn K.J.E., Lalliot M.D., Mullis P.E., Antonarakis S.E.,
RA: Kawasaki K., Asakawa S., Ito F., Shimizu N.,
RA: "Positional cloning of the APECED gene.";
RA: Nat. Genet. 17:393-398(1997).
RL: [2]
RS: SEQUENCE FROM N.A. (ISOFORM 1).
RC: TISSUE=Thymus;
RX: MEDLINE=98061087; PubMed=9398840;
RA: Aaltonen J., Björnsen P., Perheentupa J., Horelli-Kuitunen N.,
RA: Palotie A., Peltonen L., Lee Y.S., Francis F., Hennig S., Thiel C.,
RA: Lehnach H., Yaspo M.-L.;
RA: "An autoimmune disease, APECED, caused by mutations in a novel gene
featuring two PHD-type zinc-finger domains.";
RA: Nat. Genet. 17:399-403(1997).
RL: [3]
RS: SEQUENCE FROM N.A.
RA: Lee Y.S., Francis F., Hennig S., Thiel C., Reinhard R., Lehnach H.,
RA: Yaspo M.-L.;
RA: Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RL: [4]
RS: SEQUENCE FROM N.A.
RA: MEDLINE=20289799; PubMed=10830953;
RA: Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA: Paki H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA: Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA: Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA: Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA: Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA: Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA: Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,
RA: Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,
RA: Ramser J., Beck A., Klages S., Hennig S., Riesseltmann L., Dagand E.,
RA: Lehnach H., Reinhardt K., Yaspo M.-L.;
RA: "The DNA sequence of human chromosome 21.";
RA: Nature 405:311-319(2000).
RL: [5]

RA SUBCELLULAR LOCATION.
 RA MEDLINE-99135905; PubMed-9931333;
 RA Björres P., Pelto-Huikko M., Kaukonen J., Aaltonen J., Peltonen L.,
 RA Umanen I.;
 RA "Localization of the APECED protein in distinct nuclear structures.";
 RA Hum. Mol. Genet. 8:259-266(1999).
 RA [6]
 RA PARTIAL SEQUENCE, SUBUNIT STRUCTURE, DNA-BINDING, AND PHOSPHORYLATION.
 RA PubMed-11533054;
 RA Kumar P.G., Laloraya M., Wang C.-Y., Ruan Q.-G., Davoodi-Semlromi A.,
 RA Kao K.-J., She J.-X.;
 RA "The autoimmune regulator (AIRE) is a DNA-binding protein.";
 RA J. Biol. Chem. 276:41357-41364(2001).
 RA [7]
 RA SUBCELLULAR LOCATION, TRANSCRIPTION ACTIVATION, AND MUTAGENESIS OF
 RA LEU-28: CYS-302 AND CYS-437.
 RA PubMed-11274163;
 RA Pitkanen J., Vaehemurto P., Krohn K.J.E., Peterson P.;
 RA "Subcellular localization of the autoimmune regulator protein.
 RA characterization of nuclear targeting and transcriptional activation
 RA domain";
 RA J. Biol. Chem. 276:19597-19602(2001).
 RA [8]
 RA VARIANT APECED PRO-28.
 RA MEDLINE-99103466; PubMed-9888391;
 RA Heino M., Scott H.S., Chen Q., Peterson P., Maenpää U.,
 RA Papasavvas M.-P., Miltaz L., Barras C., Rossier C., Chrousos G.P.,
 RA Stratakis C.A., Nagamine K., Kudoh J., Shimizu N., Krohn K.J.E., Antonarakis S.E.,
 RA Antonarakis S.E., Krohn K.J.E.;
 RA "Mutation analyses of North American APS-1 patients.";
 RA Hum. Mutat. 13:69-74(1999).
 RA [9]
 RA VARIANT ARG-278.
 RA MEDLINE-98381937; PubMed-9717837;
 RA Scott H.S., Heino M., Peterson P., Miltaz L., Laijola M.D.,
 RA Bettele C., Cohen A., Serl M., Lerone M., Romeo G., Collin P.,
 RA Salo M., Metcalfe R., Weetman A., Papasavvas M.-P., Rossier C.,
 RA Nagamine K., Kudoh J., Shimizu N., Krohn K.J.E., Antonarakis S.E.;
 RA "Common mutations in autoimmune polyendocrinopathy-candidiasis-
 RA ectodermal dystrophy patients of different origins.";
 RA Mol. Endocrinol. 12:1112-1119(1998).
 RA [10]
 RA VARIANTS APECED LEU-15; MET-16; PRO-28; ARG-78; LEU-80;
 RA GLU-83; CYS-85; CYS-90; ARG-93; MET-301; TYR-311 AND GLN-326, AND
 RA VARIANT ARG-278.
 RA PubMed-11524731;
 RA Heino M., Peterson P., Kudoh J., Shimizu N., Antonarakis S.E.,
 RA Scott H.S., Krohn K.J.E.;
 RA "APECED mutations in the autoimmune regulator (AIRE) gene.";
 RA Hum. Mutat. 18:205-211(2001).
 RA -1- FUNCTION: Probable transcriptional regulator protein that binds to
 RA DNA as dimer and tetramer, but not as a monomer. Binds to G-
 RA tandems in an A/T-rich environment; the preferred motif is a
 RA tandem repeat of ATTGGTTA combined with a TTAATTA-box. May be
 RA involved in immune regulation.
 RA -1- SUBUNIT: Homodimer and homotetramer. Interacts with CREBP.
 RA -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic; associated with
 RA tubular structures and in discrete nuclear dots resembling ND10
 RA nuclear bodies. May shuttle between nucleus and cytoplasm.
 RA -1- ALTERNATIVE PRODUCTS: At least 3 isoforms: 1/AIRE-1 (shown here),
 RA 2/AIRE-2 and 3/AIRE-3; may be produced by alternative
 RA splicing.
 RA -1- TISSUE SPECIFICITY: Widely expressed. Expressed at higher level in
 RA thymus (medullary epithelial cells and monocyte-dendritic cells),
 RA pancreas, adrenal cortex, and testis. Expressed at lower level in
 RA the spleen, fetal liver and lymph nodes. Isoforms 2 and 3 seem to
 RA be less frequently expressed than isoform 1, if at all.
 RA -1- DOMAIN: The L-X-X-L-L repeats may be implicated in binding to
 RA nuclear receptors.
 RA -1- DOMAIN: The HSR domain is required for localization on tubular
 RA structures (N-terminal part) and for homodimerization.
 RA -1- DOMAIN: Disruption of the first PHD domain has been shown to lead
 RA to reduced transcriptional activity and to localization of the

CC protein mainly in the cytoplasm in small granules.
 CC -1- PTM: Phosphorylated. Phosphorylation could trigger
 CC oligomerization.
 CC -1- DISEASE: Defects in AIRE are a cause of autoimmune poly-
 CC endocrinopathy candidiasis ectodermal dystrophy (APECED), also
 CC called autoimmune polyglandular syndrome type I (APS-1), an
 CC autosomal recessive disease characterized by: (1) autoimmune
 CC polyendocrinopathies: hypoparathyroidism, adrenocortical failure,
 CC IDDM, gonadal failure, hypothyroidism, pernicious anemia, and
 CC hepatitis; (2) chronic mucocutaneous candidiasis; (3) ectodermal
 CC dystrophies: vitiligo, alopecia, keratopathy, dystrophy of dental
 CC enamel, nails and tympanic membranes. In addition, a high
 CC proportion of patients develop squamous cell carcinoma of the oral
 CC mucosa. The disease is reported worldwide but is exceptionally
 CC prevalent among the Finnish population (incidence 1:25000) and
 CC the Iranian Jews (incidence 1:9000).
 CC -1- SIMILARITY: CONTAINS 1 HSR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 SAND DOMAIN.
 CC -1- DATABASE: NAME-AIRE; NOTE-Article on AIRE and APECED;
 CC WWW="http://chr21.rz-berlin.mpg.de/APECED.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB006682; BAA23988.1; -
 CC EMBL; AB006683; BAA23989.1; -
 CC EMBL; AB006684; BAA23990.1; -
 CC EMBL; AB006684; BAA23991.1; -
 CC EMBL; AB006684; BAA23992.1; -
 CC EMBL; AB006685; BAA23993.1; -
 CC EMBL; Z97990; CAB10790.1; -
 CC EMBL; AJ009610; CAA08759.1; -
 CC EMBL; AP001754; BAA95560.1; -
 CC Genew; HGNC:360; AIRE.
 CC MIM: 240300; -
 CC InterPro: IPR000770; SAND_domain.
 CC InterPro: IPR004865; Sp100.
 CC InterPro: IPR001965; ZnF_PHD.
 CC Pfam: PF00628; PHD; 2.
 CC Pfam: PF01342; SAND; 1.
 CC Pfam: PF03172; Sp100; 1.
 CC SMART; SM00249; PHD; 2.
 CC SMART; SM00258; SAND; 1.
 CC PROSITE; PS00864; SAND; 1.
 CC PROSITE; PS01359; ZF_PHD_1; 2.
 CC PROSITE; PS0016; ZF_PHD_2; 1.
 CC KW Nuclear protein; DNA-binding; Transcription regulation; Zinc-finger;
 CC Repeat; Phosphorylation; Alternative splicing; Polymorphism;
 CC Disease mutation.
 CC KW
 CC DOMAIN 1 106
 CC DOMAIN 181 280
 CC ZN_FING 296 343
 CC ZN_FING 434 475
 CC DOMAIN 113 133
 CC DOMAIN 7 520
 CC REPEAT 7 11
 CC REPEAT 63 67
 CC REPEAT 414 418
 CC REPEAT 516 520
 CC VANSPLIC 1 292
 CC
 CC HSR.
 CC SAND.
 CC PHD-TYPE 1.
 CC PHD-TYPE 2.
 CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC 4 X 5 AA REPEATS OF L-X-X-L-L.
 CC 1.
 CC 2.
 CC 3.
 CC 4.
 CC MATDAALRLRLRLHRTETAVAVDSAFPLLLHALADHDVVPED
 CC KFOETLHLKEKGCQAFHALSWLTLDDSTAIIDFWRYLF
 CC KDYNLERYGRQPIIDSPKVDLSQPKRGKPPAPVAPVAF
 CC PPRPLPTKRKASEARAAPALPFRGTAASVAGSOLKAPPK
 CC KPSSAEOGRPLDGLGIGTOMASVORAVMSGSDPGARCA
 CC VEGTILIOVFSSGSKKCIQVGGERTTSKEDSDSGNRKA
 CC RSSSGPKRLVAKAGQGAAPGGEARLQDQGSVPAPLAPS
 CC DPQLH -> MMLVYSSGAPGTQGPARNRVFPFGMAPGVC

Query Match 100.0%; Score 2902; DB 1; Length 545;
 Best Local Similarity 100.0%; Pred. No. 3.3e-152;
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATDAALRLRLHRTETAVAVDSAPFLALADHDVVPEDKQETLHLEKEBCCPOAFH 60
 |||||||
 DB 1 MATDAALRLRLHRTETAVAVDSAPFLALADHDVVPEDKQETLHLEKEBCCPOAFH 60
 |||||||

QY 61 ALLSWLITQDSTAILDFWHLFEKDYNIERYGRLOPILDSFPKVDLSQPKRGKRPVVK 120
 |||||||
 DB 61 ALLSWLITQDSTAILDFWHLFEKDYNIERYGRLOPILDSFPKVDLSQPKRGKRPVVK 120
 |||||||

QY 121 ALVPPRLPTKKASEEAAAPALTPRGTSFGSGLAKRPPKPKRESSAQOQLPLGNG 180
 |||||||
 DB 121 ALVPPRLPTKKASEEAAAPALTPRGTSFGSGLAKRPPKPKRESSAQOQLPLGNG 180
 |||||||

QY 181 IOTMSAVQRAVAMSSGDVPGARGAVEGILIOVFESGSKCIQYGVGEYTPSKFEDSG 240
 |||||||
 DB 181 IOTMSAVQRAVAMSSGDVPGARGAVEGILIOVFESGSKCIQYGVGEYTPSKFEDSG 240
 |||||||

QY 241 SGKNKARSSGPKPLVRAKAGAGAPGGEGARLGOOGSVAPALPSPDQLHOKNDECA 300
 |||||||
 DB 241 SGKNKARSSGPKPLVRAKAGAGAPGGEGARLGOOGSVAPALPSPDQLHOKNDECA 300
 |||||||

QY 301 VCRGGELICDGCGRATFLACLSPLREIPSGTWKSSCLQATVOEVOVPRAEPRPOEP 360
 |||||||
 DB 301 VCRGGELICDGCGRATFLACLSPLREIPSGTWKSSCLQATVOEVOVPRAEPRPOEP 360
 |||||||

QY 361 PVETPLPPLGKSAGEVGPPEGLAGMDTLYVKHLRPPSAAPLPGIDSSALHPLLCY 420
 |||||||
 DB 361 PVETPLPPLGKSAGEVGPPEGLAGMDTLYVKHLRPPSAAPLPGIDSSALHPLLCY 420
 |||||||

QY 421 GPEGQONLAPGARGCVDGTDVLRCTCAAFHNRCHPAGTSRPGTGLRCSGSDYT 480
 |||||||
 DB 421 GPEGQONLAPGARGCVDGTDVLRCTCAAFHNRCHPAGTSRPGTGLRCSGSDYT 480
 |||||||

QY 481 PAPVGVLAAPPARLAPAPADDTASHEPALHRDLESLSHETEDGILQWAIQSNARPA 540
 |||||||
 DB 481 PAPVGVLAAPPARLAPAPADDTASHEPALHRDLESLSHETEDGILQWAIQSNARPA 540
 |||||||

QY 541 APFPS 545
 |||||
 DB 541 APFPS 545

RESULT 2
 AIRE_MOUSE
 ID AIRE_MOUSE STANDARD; PRT: 552 AA.
 AC 09J0B3; 09J1X0; 09J1M1; 09J1M2; 09J1M3; 09J1M4; 09J1M5; 09J1M6;
 AC 09J1M7; 09J1M8; 09J1M9; 09J1M0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Autoimmune regulator (Autoimmune polyendocrinopathy candidiasis
 ectodermal dystrophy protein homolog) (APECED protein homolog).
 GN AIRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1A).
 RA Pubmed=10593569;
 RA Shi J.-D., Wang C.-Y., Marron M.P., Ruan Q.-G., Huang Y.Q.,
 RT Deleter J.C., She J.-X.;
 RT "Chromosomal localization and complete genomic sequence of the murine
 RT autoimmune regulator gene (Aire).";
 RL Autoimmunity 31:47-53(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1A).
 RC STRAIN=129/SV;
 RX MEDLINE=99160890; PubMed=10049735;

RA Miltaz L., Rossier C., Helmo M., Petersen P., Krohn K.J.E., Gos A.,
 RA Morris M.A., Kudoh J., Shimizu N., Antonarakis S.E., Scott H.S.;
 RT "Isolation and characterization of the mouse Aire gene.";
 RL Biochem. Biophys. Res. Commun. 255:483-490(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1A).
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RX MEDLINE=99168902; PubMed=10049587;
 RA Wang C.-Y., Shi J.-D., Davoodi-Semlromi A., She J.-X.;
 RT "Cloning of Aire, the mouse homologue of the autoimmune regulator
 RT (AIRE) gene responsible for autoimmune polyglandular syndrome type 1
 RT (ASPI).";
 RL Genomics 55:322-326(1999).
 RN [4]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=129;
 RX Pubmed=10022980;
 RA Blechschmidt K., Schweiger M., Wertz K., Poulsen R.,
 RA Christensen H.-M., Rosenthal A., Lehrach H., Yaspo M.-L.;
 RT "The mouse Aire gene: comparative genomic sequencing, gene
 RT organization, and expression.";
 RL Genome Res. 9:158-166(1999).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 1A; 1B; 1C; 1D; 2A; 2B; 2C; 2D; 3A; 3B;
 RP 3C AND 3D).
 RC STRAIN=C57BL/6, NOD, and STL;
 RX Pubmed=10550218;
 RA Ruan Q.-G., Wang C.-Y., Shi J.-D., She J.-X.;
 RT "Expression and alternative splicing of the mouse autoimmune regulator
 RT gene (Aire).";
 RL J. Autoimmun. 13:307-313(1999).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 1A), AND SUBCELLULAR LOCATION.
 RC TISSUE=Kidney, and Thymus;
 RX Pubmed=11156688;
 RA Halonen M., Peltto-Huikko M., Eskelin P., Peltonen L., Ullmanen I.,
 RA Kolmer M.;
 RT "Subcellular location and expression pattern of autoimmune regulator
 RT (Aire), the mouse orthologue for human gene defective in autoimmune
 RT polyendocrinopathy candidiasis ectodermal dystrophy (APECED).";
 RL J. Histochem. Cytochem. 49:197-208(2001).
 RN [7]
 RP SUBUNIT STRUCTURE, AND PHOSPHORYLATION.
 RC Pubmed=11533054;
 RA Kumar P.G., Laloraya M., Wang C.-Y., Ruan Q.-G., Davoodi-Semlromi A.,
 RA Kao K.-Y., She J.-X.;
 RT "The autoimmune regulator (AIRE) is a DNA-binding protein.";
 RL J. Biol. Chem. 276:41357-41364(2001).
 CC -1- FUNCTION: Probable transcriptional regulator protein that binds to
 CC DNA as dimer and tetramer, but not as a monomer. Binds to G-
 CC tandem repeat in an A/T-rich environment; the preferred motif is a
 CC involved in immune regulation (By similarity).
 CC -1- SUBUNIT: Homodimers and homotetramers.
 CC -1- SUBCELLULAR LOCATION: Predominantly nuclear but also cytoplasmic.
 CC Found in nuclear body-like structures and in a filamentous
 CC vimentin-like pattern.
 CC -1- ALTERNATIVE PRODUCTS: At least 12 isoforms; 1a (shown here), 1b,
 CC 1c, 1d, 2a, 2b, 2c, 2d, 3a, 3b, 3c and 3d; are produced by
 CC alternative splicing. Isoforms 3a to 3d are probably
 CC nonfunctional.
 CC -1- TISSUE SPECIFICITY: Expression may be restricted to a small number
 CC of scattered cells in most tissues. Highly expressed in a few
 CC cells in the medulla of the thymus. Detected at very low levels in
 CC thymus, lymph node, liver, brain, ovary, lung, testis, kidney,
 CC heart, spleen, bone marrow, skeletal muscle and adrenal gland.
 CC Isoforms 1a to 1d predominate, isoforms 2a to 2d are intermediate
 CC and isoforms 3a to 3d are expressed at extremely low levels.
 CC -1- DOMAIN: The L-X-X-L-L repeats may be implicated in binding to
 CC nuclear receptors.
 CC -1- DOMAIN: The N-terminal HSR domain is required for localization
 CC on tubular structures (By similarity).
 CC -1- PTM: Phosphorylated.

CC	-1	SIMILARITY:	CONTAINS 1 HSR DOMAIN.
CC	-1	SIMILARITY:	CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC	-1	SIMILARITY:	CONTAINS 1 SAND DOMAIN.
CC			-----
CC			This SWISS-PROT entry is copyright. It is produced through a collaboration
CC			between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC			the European Bioinformatics Institute. There are no restrictions on its
CC			use by non-profit institutions as long as its content is in no way
CC			modified and this statement is not removed. Usage by and for commercial
CC			entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC			or send an email to license@isb-sib.ch).
CC			-----
DR	EMBL:	AFI05002;	AAD64621.1; -
DR	EMBL:	AFI28772;	AAF36481.1; -
DR	EMBL:	AFI28773;	AAF36482.1; -
DR	EMBL:	AJ007715;	CAN07620.1; -
DR	EMBL:	AE079536;	AAD20444.1; -
DR	EMBL:	AJI32243;	CAB36460.1; -
DR	EMBL:	AFI28115;	AAF36460.1; -
DR	EMBL:	AFI28116;	AAF36461.1; -
DR	EMBL:	AFI28117;	AAF36462.1; -
DR	EMBL:	AFI28118;	AAF36463.1; -
DR	EMBL:	AFI28119;	AAF36464.1; -
DR	EMBL:	AFI28120;	AAF36465.1; -
DR	EMBL:	AFI28121;	AAF36466.1; -
DR	EMBL:	AFI28122;	AAF36467.1; -
DR	EMBL:	AFI28123;	AAF36468.1; -
DR	EMBL:	AFI28124;	AAF36469.1; -
DR	EMBL:	AFI28125;	AAF36470.1; -
DR	EMBL:	AJ243821;	CAB6141.1; -
DR	MGI:	MGI:1338803;	Aire.
DR	InterPro:	IPR000770;	SAND_domain.
DR	InterPro:	IPR004865;	SP100.
DR	InterPro:	IPR001965;	Znf_PHD.
DR	pfam:	PF00628;	PHD_2.
DR	pfam:	PF01345;	SAND; 1.
DR	pfam:	PF03172;	SP100; 1.
DR	SMART:	SM00249;	PHD; 2.
DR	SMART:	SM00258;	SAND; 1.
DR	PROSITE:	PS50864;	SAND; 1.
DR	PROSITE:	PS01359;	ZF_PHD_1; 2.
DR	PROSITE:	PS50016;	ZF_PHD_2; 1.
KW	Nuclear protein; DNA-binding; transcription regulation; zinc-finger; Repeat; Phosphorylation; Alternative splicing.		
FT	DOMAIN	1	107
FT	DOMAIN	182	282
FT	ZN_FING	298	345
FT	ZN_FING	434	475
FT	DOMAIN	114	134
FT	DOMAIN	8	524
FT	REPEAT	8	12
FT	REPEAT	64	68
FT	REPEAT	414	418
FT	REPEAT	520	524
FT	VARSPLIC	265	268
FT			MISSING (IN ISOFORM 1C, ISOFORM 1D, ISOFORM 2C, ISOFORM 2D, ISOFORM 3C AND ISOFORM 3D).
FT	VARSPLIC	236	236
FT			MISSING (IN ISOFORM 1B, ISOFORM 1D, ISOFORM 2B, ISOFORM 3B AND ISOFORM 3D).
FT	VARSPLIC	367	425
FT			MISSING (IN ISOFORM 2A, ISOFORM 2B, ISOFORM 2C AND ISOFORM 2D).
FT	VARSPLIC	368	409
FT			IIVGLSASERTKRPSELSDAAYTVVNLAPPAFL L -> DOSPLQILGRIDSLRHRTGSCSTHLPSPSTMACO GGRLIC (IN ISOFORM 3A, ISOFORM 3B, ISOFORM 3C AND ISOFORM 3D).
FT	VARSPLIC	410	552
FT			MISSING (IN ISOFORM 3A, ISOFORM 3B, ISOFORM 3C AND ISOFORM 3D).
SO	SEQUENCE	552 AA:	59042 MW; BF30F8f66b71239A CRC64;
Query Match		69.1%;	Score 2005; DB 1; Length 552;
Best Local Similarity		71.8%;	Pred. No. 4e-103;
Matches 395; Conservative		37;	Mismatches 108; Indels 10; Gaps 6;

QY	4	DAALRRLRLHRLHRELAIVANDSAFPLLHIALHADHVVPEKFOETLHKKKEGPOAFHALL	63
Dd	5	DGMLRLHLHRLHRELAIVANDSAFPLHIALHADHVVPEDFQETLHKKKEGPOAFHALL	64
QY	64	SWLLTQDSTAILDFWFLVKDYNLEBYGRLOPLDPSFKVDLSQPRKGRPPAPKALV	12
Dd	65	SWLLTRDSCAILDFWRLLEFKDYNLEBYSLHSLIDGFPKVDLNSRRKGRPLAGPKAAV	12
QY	124	PPPRPLPTKKAASEBARAAAPALLTRGRASPGSOLKAPPKKPRESSAQORLPLGNGIQT	18
Dd	125	LPBRPPTKKAALKEBRATPPATLASKSVSSPGSHLKTTPPKKPDGNLESQHLPLGNGIQT	18
QY	184	MSASVORAAVMSGVPVPGARGAVEILLIOQVVEESGSKKCIQVGEFEPTSKFED -SSG	24
Dd	185	MAASVORAVTAVASGDVPTGRGAVEILLIOQVFEESRSKKCIQVGEFEPTPKFPDPSGNT	24
QY	243	KNKRASSSGPKPLVBAKGOAGAAPRGGEARLQOQGSVPAPLALPSPDPLHOKNDECAVC	30
Dd	245	KNKRASSGSLKRYVBAKGOAYTIPERDDQXKYGQCGVPRPLSLSPREVONKNDECAVC	30
QY	303	RGGELICCDGCPRAFHILACSLPPLREIIPSGTWRCSSCLQATVOEYOPRAEPPROEPPV	36
Dd	305	HGGELICCDGCPRAFHILACSLPPLQEIIPSGILMRCSCLOGRVQONLSQPEVSRPELPA	36
QY	363	ETPLRPGLSRSGEEVSRPGPEGLAGMOTTYLKKHLPARPSAAPPLGDDSSMLHLCLVCP	42
Dd	365	ETPLRVGLRSRSEKTRGSRRELKSSDAAVYVYNLVAHPAAPL -LEPSMLICLLSAGN	42
QY	423	EGQONLAPGARGVCGDSTDVLRCTHCAAAHMHRCHPAGTSRPGTGRJSCSGSDVTPA	48
Dd	423	EGRPGPAPRSACSVCGDTEVLRCAHCAAAHMHCHPPTAAARGCTNLRCCKSCADSTPT	48
QY	483	P-VESVLPAR-SPARLAPGPAA--DDTASHEPALHRDLESLSEHTPDGILQMAIOSMAR	538
Dd	483	PCTPEAVPTSGPRAPGLAKVGDSDASHDEVTLHRDLESLSEHNSFDGILQMAIOSMSR	542
QY	539	PAA-- -PPPS 545	
Dd	543	PLAETPPRESS 552	

ID	CHD3_HUMAN	STANDARD:	PRT: 1944 AA.
AC	Q12873:		
DT	15-JUL-1999 (Rel. 38, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Chromodomain helicase-DNA-binding protein 3 (CHD-3) (Mi-2 autoantigen		
DE	240 kDa protein) (M12-alpha).		
GN	CHD3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Fetal;		
RA	MEDLINE=97470991; Pubmed=9326634;		
RA	Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.,		
RT	"Characterization of the CHD family of proteins."		
RL	Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).		
RN	[2]		
RP	SEQUENCE OF 121-654 FROM N.A.		
RC	TISSUE=Thymus;		
RA	MEDLINE=96013633; Pubmed=7560064;		
RA	Ge Q., Nitsasena D.S., O'Brien C.A., Frank M.B., Targoff I.N.;		
RT	"Molecular analysis of a major antigenic region of the 240 kD protein		
RL	of Mi-2 autoantigen."		
RL	J. Clin. Invest. 96:1730-1737(1995).		
CC	-1- FUNCTION: POSSIBLE TRANSCRIPTION ACTIVATOR (BY SIMILARITY).		
CC	-1- SUBUNIT: Central component of the nucleosome remodelling and		
CC	histone deacetylase (NuRD) complex.		


```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: ONE OF THE MAIN ANTIGENS REACTING WITH ANTI-MI-2 POSITIVE
CC SERA OF DERMATOMYOSITIS.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF006515; AAB87383.1; -
DR EMBL: U08379; AAC50228.1; -
DR Genew: HGNC:1918; CHD3.
DR MIM: 602120; -
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAH_box.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; Znf_PHD.
DR Pfam: PF00176; SNF2_N; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00385; chromo; 1.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00298; CHROMO; 2.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR PROSITE: PS00598; CHROMO_2; 1.
DR PROSITE: PS50013; CHROMO_2; 2.
DR PROSITE: PS00690; DEAH-ATP-HELICASE; 1.
DR PROSITE: PS01359; ZF_PHD_1; 2.
DR PROSITE: PS50016; ZF_PHD_2; 2.
KW Chromatin regulator; DNA-binding; ATP-binding; Helicase;
KW Nuclear protein; Repeat; Transcription regulation; Activator; Antigen;
KW Zinc-finger.
FT ZN_FING 379 426 PHD-TYPE 1.
FT ZN_FING 456 503 PHD-TYPE 2.
FT DOMAIN 494 594 CHROMO 1.
FT DOMAIN 631 673 CHROMO 2.
FT NP_BIND 761 768 ATP (POTENTIAL).
FT SITE 883 886 DEAH_BOX.
FT DOMAIN 206 221 POLY-ALA.
FT DOMAIN 243 246 POLY-PRO.
FT DOMAIN 355 358 POLY-LYS.
FT DOMAIN 434 446 POLY-GLU.
FT DOMAIN 697 703 POLY-LYS.
FT CONFLICT 121 126 GEGDGG->PHFOOK (IN REF. 2).
FT CONFLICT 309 312 MISSING (IN REF. 2).
FT CONFLICT 653 653 W->G (IN REF. 2).
SQ SEQUENCE 1944 AA; 220691 MW; 071266F1D5FD7D335 CRC64;

Query Match 9.3%; Score 270.5; DB 1; Length 1944;
Best Local Similarity 23.2%; Pred. No. 1.1e-07;
Matches 102; Conservative 46; Mismatches 154; Indels 137; Gaps 17;

```

```

DB 310 VFQSDGEPPEAESDLDGSHSASGRPDGPFRTKLRGRGRKKKYLQ----CPAV 365
QY 283 LALPSDPLHOKNEDECAVCPRDGEILCCDGPRAFLACLSPLEIPSGTRCSCLO 342
DB 366 AGESEVDSYENDHOCYCEVCOGGELIICDTCPRAHLVCLDELDRAIPGKSCPCXK 425
QY 343 ATVOEVOGRAEPRPOEPPVETPLPPGLRSAGEEVNPGPEPLAGMDTLVYKHLPAAPS 402
DB 426 EGVOMEAKEEEDEVEE-----GEE-EGEKEE-----DHMEY----- 458
QY 403 AAPPLGDSALHPLLCVGPBEGQNLAPARGCVCGDGTDLVRLCTCAAPFHRCHFPAG 462
DB 459 -----CRVCKDGGELLCDDACISSYHTHCLNPL 487
QY 463 TSPRGTLRCRSGDVTAPAVEGLV-----APSPARLAPAKODTASHEPALH 512
DB 488 PDINGEWLRCRCICPVLKGRQKILHWRGEPVAVPAPQADGNP---DVPPRPLOG 544
QY 513 RDLLESLSEHTFDGIIQW 531
DB 545 RSEREFVK---WGLSYW 560

RESULT 4
CHD4_HUMAN STANDARD; PRT; 1912 AA.
ID CHD4_HUMAN
AC Q14839;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromodomain helicase-DNA-binding protein 4 (CHD-4) (MI-2 autoantigen
DE 218 kDa protein) (MI2-beta).
GN CHD4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6017437; Pubmed=7575689;
RA Seelig H.P., Moosbrugger T., Ehrfeld H., Fink T., Renz M., Genth E.;
RT "The major dermatomyositis specific MI-2 autoantigen is a presumed
RT helicase involved in transcriptional activation."
RL Arthritis Rheum. 38:1389-1399(1995).
CC -1- FUNCTION: POSSIBLE TRANSCRIPTION ACTIVATOR (BY SIMILARITY).
CC -1- SUBUNIT: Central component of the nucleosome remodelling and
CC histone deacetylase (NURD) complex.
CC -1- DISEASE: ONE OF THE MAIN ANTIGENS REACTING WITH ANTI-MI-2 POSITIVE
CC SERA OF DERMATOMYOSITIS.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X86691; CAA60384.1; -
DR Genew: HGNC:1919; CHD4.
DR MIM: 603277; -
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAH_box.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_RING.
DR Pfam: PF00176; SNF2_N; 1.

```

DR Pfam: PF00271; helicase-C; 1.
 DR Pfam: PF00385; chromo; 1.
 DR Pfam: PF00628; PHD; 2.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR PROSITE: PS00598; CHROMO_1; 2.
 DR PROSITE: PS00013; CHROMO_2; 2.
 DR PROSITE: PS00690; DEAH_ATP_HELICASE; 1.
 DR PROSITE: PS00359; ZF_PHD_1; 2.
 DR PROSITE: PS00016; ZF_PHD_2; 2.
 DR Chromatin regulator; DNA-binding; ATP-binding; Helicase;
 KW Nuclear protein; Repeat; Transcription regulation; Activator; Antigen;
 ZINC-finger
 FT ZN_FING 370 417 PHD-TYPE 1.
 FT ZN_FING 449 496 PHD-TYPE 2.
 FT DOMAIN 494 594 CHROMO 1.
 FT DOMAIN 622 697 CHROMO 2.
 FT NP_BIND 751 758 ATP (POTENTIAL).
 FT SITE 873 876 DEAH_BOX.
 FT DOMAIN 50 59 POLY-LYS.
 FT DOMAIN 94 98 POLY-GLU.
 FT DOMAIN 114 119 POLY-LYS.
 FT DOMAIN 134 138 POLY-GLU.
 FT DOMAIN 139 144 POLY-ASP.
 FT DOMAIN 227 235 POLY-ALA.
 FT DOMAIN 248 252 POLY-PRO.
 FT DOMAIN 350 354 POLY-LYS.
 FT DOMAIN 1052 1055 POLY-LEU.
 FT DOMAIN 1294 1301 POLY-GLU.
 FT DOMAIN 1665 1668 POLY-GLU.
 SQ SEQUENCE 1912 AA: 217989 MW: 80A73300E80166CE CRC64;

Query Match 8.3%; Score 241; DB 1; Length 1912;
 Best Local Similarly 23.1%; Pred. No. 4.4e-06;
 Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14;

QY 129 PTRKASEEARAAAPALTPRGTSAPGSQLAKAPRKRESSAEQORPLPGNGIOTMSASV 188
 Db 217 PFKSSGSAVAAMAAVAVSWMTATEV-APPPVEVPIRKAKTRGKGPMA----- 270
 QY 189 QRAVAMSSGDVPGARVAGVEGILIOVEPSG-----SKKCIQVGFYPSKPED----- 238
 Db 271 -RRPKSPRPVPAKKPKKVPAPLKITKIFGFGSKRRKSSSEDDLDVSDPDASINSY 329
 QY 239 SGSGKNAARSSSGPKPLVRAKGAAGAAFGGGEARLGQGSVPAPLALPSDQLHOKNEDE 298
 Db 330 SVSGGTSRSRSRSRKRKLRTRK-----KKRKGEE-EVTAVDGYETD-----HODY 372
 QY 299 CAVCRDGEILICDCPRRAFLACTSPPLREIPSGTWRCSSCLQATVQEVQPRAEPRRQ 358
 Db 373 CEVCGQGGELITLCTCRATHWCLDPMKAPGKMSCPCEK---EGIQWEAKED--- 426
 QY 359 EPPVETPLPGLSAGEEVGPPPEPLAGMDTLVYKHLPAAPSAAPPLGLDSSALHPL 418
 Db 427 -----NSEGEILIEVGGDLEEDD-----HHMEF----- 451
 QY 419 CVGPEGOQNLAPGARCVGDDGVDVLRCTICAAAFHWRCHFPAGTSRPGTLKRCSCSD 478
 Db 452 -----CVCKCKDGGELILCCDTCPSSTYHICLNPPLPEIPNGEMLCPRTCP 496
 QY 479 VTPAPVEGVL-----APSPARL-APPAKDDTASHPEPALHRDLESLSHTEHTDGILOW 531
 Db 497 ALKGKGVOKIILKMWGPPSPFPVPRPPDADPNTPSPKPLLEGRRPDRPFVK---WQGMASYW 553

RESULT 5
 SP11_HUMAN
 ID SP11_HUMAN STANDARD: PRT; 689 AA.
 AC Q9H58; Q9H58; Q9H58; Q14976; Q14977;
 DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sp110 nuclear body protein (Speckled 110 kDa) (Transcriptional
 DE coactivator Sp110) (Interferon-induced protein 41/75).
 GN SP110.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4), AND PHOSPHORYLATION (ISOFORM
 RP 2).
 RC TISSUE=Lymphoma;
 RA MEDLINE=94043285; PubMed=7693701;
 RT Kaderoff S., Gewert D.R., Galabru J., Hovanessian A.G., Meurs E.F.;
 RT "Molecular cloning of two new interferon-induced, highly related
 RT nuclear phosphoproteins.";
 RL J. Biol. Chem. 268:24432-24441(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND SUBCELLULAR LOCATION.
 RC TISSUE=Spleen;
 RA MEDLINE=20372745; PubMed=10913195;
 RL Bloch D.B., Nakajima A., Gulick T., Chliche J.-D., Orch D.,
 RL de la Monte S.M., Bloch K.D.;
 RT "Sp110 localizes to the PML-Sp100 nuclear body and may function as a
 RT nuclear hormone receptor transcriptional coactivator.";
 RL Mol. Cell. Biol. 20:6138-6146(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 5).
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transcription factor. May be a nuclear hormone receptor
 CC coactivator. Enhances transcription of genes with retinoic acid
 CC response elements (RARE).
 CC -1- SUBCELLULAR LOCATION: Nuclear; found in the nuclear body.
 CC -1- ALTERNATIVE PRODUCTS: 5 isoforms; 1 (shown here), 2/FT5/75,
 CC 3/Sp110b, 4/FT14/41 and 5; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in peripheral blood
 CC leukocytes and spleen. Detected at intermediate levels in thymus,
 CC prostate, testis, ovary, small intestine and colon, and at low
 CC levels in heart, brain, placenta, lung, liver, skeletal muscle,
 CC kidney and pancreas.
 CC -1- INDUCTION: By interferon gamma and by all-trans retinoic acid.
 CC -1- PTM: Phosphorylated (isoform 2).
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 HSR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -1- SIMILARITY: CONTAINS 1 SAND DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L22342; AAA18806.1; -
 CC EMBL: L22343; AAD13402.1; -
 CC EMBL: AF280094; AAF99318.1; -
 CC EMBL: AF280095; AAG09826.1; -
 CC EMBL: BC019059; AAH19059.1; -
 CC Genew: HGNC:5401; SP110.
 CC MIM: 604457; -
 CC InterPro: IPR001487; Bromodomain.
 CC InterPro: IPR000770; SAND_domain.
 CC InterPro: IPR004865; SP100.
 CC InterPro: IPR001965; Znf_PHD.
 CC Pfam: PF00628; PHD; 1.
 CC Pfam: PF01342; SAND; 1.
 CC Pfam: PF03172; SP100; 1.
 CC SMART: SM00297; BROMO; 1.

DR SMART: SM00249; PHD: 1.
 DR SMART: SM00258; SAND: 1.
 DR PROSITE: PS50864; SAND: 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 DR PROSITE: PS50633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS50014; BROMODOMAIN_2; FALSE_NEG.
 DR Nuclear protein; DNA-binding; Trans-acting factor; Zinc-finger;
 KW Phosphorylation; Bromodomain; Alternative splicing.
 FT DOMAIN 6 109 HSR.
 FT DOMAIN 281 294 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 428 444 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 454 535 SAND.
 FT DOMAIN 525 529 NUCLEAR HORMONE RECEPTOR INTERACTION
 (POTENTIAL).
 FT ZN_FING 534 580 PHD-TYPE.
 FT DOMAIN 581 676 BROMODOMAIN.
 FT VARSPPLIC 1 251 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 1 203 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 141 143 HHR -> PPO (IN ISOFORM 5).
 FT VARSPPLIC 252 275 IRDNPEDPEEPQEVSTPSDK -> MASSGVKNTPRWR
 RKAPGREKE (IN ISOFORM 4).
 FT VARSPPLIC 300 349 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 300 303 GTAS -> AL (IN ISOFORM 5).
 FT VARSPPLIC 531 549 RKNSECEVCCGGQLCC -> SCGLCPPTINKRELNS
 K (IN ISOFORM 4 AND ISOFORM 5).
 FT VARSPPLIC 531 539 RKNSECEV -> SGLCVLOE (IN ISOFORM 3).
 FT VARSPPLIC 540 689 MISSING (IN ISOFORM 3).
 FT VARSPPLIC 550 689 MISSING (IN ISOFORM 4 AND ISOFORM 5).
 FT VARSPPLIC 606 611 IRDGE -> NVSSSS (IN ISOFORM 2).
 FT VARSPPLIC 612 689 MISSING (IN ISOFORM 2).
 FT CONFLICT 167 167 T -> D (IN REF. 3).
 FT CONFLICT 367 367 T -> M (IN REF. 1); AAA18806).
 FT CONFLICT 464 464 L -> S (IN REF. 1; AAA18806).
 FT CONFLICT 523 523 M -> T (IN REF. 2).
 SO SEQUENCE 689 AA; 78504 MW; EBBE023101C2E59B CRC64;

Query Match 7.9%; Score 228; DB 1; Length 689;
 Best Local Similarity 19.5%; Pred. No. 8.2e-06;
 Matches 119; Conservative 44; Mismatches 138; Indels 308; Gaps 20;

2 AYDAALRLRLHREIIVAVDSAPRLHALADHDVVPEDKPEOTLHLKKEG----- 55
 7 AMEALLOHFNHOKIGIYAIHKPPFFGLDLSITTKRMESL-----EACRNLIPIV 61
 56 POAEHALSWLTDSTALDFWVLFKYNLERYGRLOPIIDSPKVDLSOPRKGR-- 113
 62 SRVYHNILTOL--ERTFNLILVLFESQINLREYNLVLTYSNF-KRVGASVYERQGRDT 117
 114 -----KP-----PAVPKALVP----- 124
 118 PILLEAPTLAEGSSLHRLPLRLPHKRPPOSCPCAPRVSEPTSSQSTELISESPSPS 177
 125 .PRPLP-----TKRKASEEARAARPAALT----- 147
 178 DPVLPRLALIOBGRSTVTKDKLTKMAEESSEMPSLITVVOASDMLIPOIRKED 237
 148 ----- 147
 238 POEMPHSLGSMPEIRDNSPEENDEPEQEVSTPSDKKKKKRRCIMSTPKRRHKKSL 297
 148 PRGASP--GSQLAK-----PKRPRESSA----- 170
 298 PRGASPHGQKLRVDYQKDKDTCNSTVETRAOKARTECARKSRSEELIDTSE 357
 171 -----EORLPL-----GNGIO-----TMSASYORA-- 191
 358 MNECKRSQKTPSTRPVYOGAASPGHGIOEKLOVDKYORKDSTWNSVMMARVQAPART 417
 192 -----VAMSSGDVPGARGAVEGLILO 212
 418 KCAKRSRKEKKKKEDICSSSKRRFQKNIHRGKPKSPDYDFGCKLPIVCGEAKGILYK 477
 213 QVESGSGKKCIQ-VGGEFTYTSKFEDSDSGKNKASSSGPKPLVLRKAKGQGAAPGGEA 271

Db 478 KKKHSSSVKCIINEDGTWLTPEFVEGKGRN-----AKMKRNIRCG-----M 523
 Qy 272 RLGGQGSVPAPLALPDPDLOHKNEDCAVCRDGEIICDCGCPRAFLACLSPLEIRIP 331
 Db 524 TLGE-----LLKRNSECEVCCGGQLCCGTCGRVFNEDCHIPPV-EAK 568
 Qy 332 SCTWRCCSC 340
 Db 569 RMLWSTFC 577
 RESULT 6
 ID TF1A_HUMAN STANDARD; PRT; 1050 AA.
 AC 015164; 095854;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription Intermediary factor 1-alpha (Tf1-1-alpha).
 GN TF1A OR TF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE=breast cancer;
 RX MEDLINE=97277352; PubMed=9115274;
 RA Theriot S., Henriquet C., Rochefort H., Cavaillès V.;
 RT "Differential interaction of nuclear receptors with the putative human
 RL transcriptional coactivator hTf1";
 RN J. Biol. Chem. 272:12062-12068(1997).
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=99144725; PubMed=10022127;
 RA Venturini L., You J., Stadler M., Gallen R., Lallemand V.,
 RA Koken H.H., Mattei M.-G., Ganser A., Chambon P., Losson R.,
 RA De The H.;
 RT "Tf1gamma, a novel member of the transcriptional intermediary factor
 RL 1 family";
 RN Oncogene 18:1209-1217(1999).
 RP [3]
 RP SEQUENCE OF 477-510 (LONG ISOFORM).
 RC TISSUE=breast cancer;
 RA Cavaillès V.;
 RL Submitted (JAN-1999) to the SWISS-PROT data bank.
 CC -!- FUNCTION: INTERACTS SELECTIVELY IN VITRO WITH THE AR2-ACTIVATING
 CC DOMAIN OF THE ESTROGEN RECEPTORS. ASSOCIATION WITH DNA-BOUND
 CC ESTROGEN RECEPTORS REQUIRES THE PRESENCE OF ESTRADIOL.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -!- SIMILARITY: CONTAINS 2 B BOX-TYPE ZINC FINGERS.
 CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF009353; AAB63585.1; -;
 DR EMBL: AF119042; AAD17258.1; -;
 DR HSSP: P29590; 1BOR.
 DR TRNSPAC: T04945; -;
 DR Genew: HGNC:11812; TF1.
 DR MIM: 603406; -;
 DR InterPro: IPR003649; Bbox_C.

DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR000315; znf_Box.
 DR InterPro: IPR001965; znf_PHD.
 DR InterPro: IPR001841; znf_Ring.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00643; zf-B_Box; 2.
 DR PRINTS: PR01406; BBOXZNFINGER.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00502; BBC; 1.
 DR SMART: SM00336; BBOX; 2.
 DR SMART: SM00297; BBOX; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS00119; ZF_BOX; 2.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS00016; ZF_PHD_2; 1.
 DR PROSITE: PS00518; ZF_RING_1; 1.
 DR PROSITE: PS00089; ZF_RING_2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Bromodomain;
 DR Zinc-finger; Alternative splicing; Nuclear protein; Coiled coil;
 DR Repeat.
 KW DOMAIN.
 FT 9 15 POLY-ALA.
 FT 56 82 RING-TYPE.
 FT 158 211 B BOX-TYPE 1.
 FT 218 259 B BOX-TYPE 2.
 FT 289 359 COILED COIL (POTENTIAL).
 FT 344 347 POLY-GLN.
 FT 754 347 NUCLEAR RECEPTOR BINDING SITE (NRBS).
 FT 826 873 PHD-TYPE.
 FT 891 907 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT 932 987 BROMODOMAIN.
 FT 477 510 MISSING (IN SHORT ISOFORM).
 FT 14 20 AASMAAS -> RIGCAP (IN REF. 1).
 FT 24 28 GSPVSG -> RGG (IN REF. 1).
 FT 109 114 SAAPS -> RGG (IN REF. 1).
 FT 350 350 D -> T (IN REF. 1).
 FT 600 600 D -> I (IN REF. 1).
 FT 608 608 M -> I (IN REF. 1).
 FT 967 967 A -> R (IN REF. 1).
 SO SEQUENCE 1050 AA; 116831 MW; D341E8022AAC67E CRC64;

Query Match 7.8%; Score 227; DB 1; Length 1050;
 Best Local Similarity 24.6%; Pred No. 1.4e-05;
 Matches 77; Conservative 46; Mismatches 104; Indels 86; Gaps 10;

71 SRAIDFMVLEKRDYLERGLQPIIDSEPKVDLSQPRKGRKPPAVPALVPPRLPT 130
 631 STIMLD-NIVKRDITID-HGQPRPPSNRTVQSPNSVPSGLAGPYTMTSVHPPIRSPS 687
 131 KRKASEEARAAAPALTPPGTASPG-----SOLKAPPKK-----PESSA 170
 688 --ASSVSGSSGSSSSKPKGADSTHKVPVLMLEPIRKQENSPPENYDEPVIVKQESD 745
 171 EQQLPLNGIQTMSASVQRAVAMSS-----GDVPGARGAVEILLIQVESSGSKK 222
 746 EESRPNANYPRIISLTLSLSSQSTSEETVLRSDAPSTGQPGEL----- 792
 223 CIYOGGEYTTSPKFEKDSGSKKARSSSGPKPLVRAKGAQGAAPGGEARLGGQSVAP 282
 793 -----HODNSSNGKSEMDPSQKSPYL-----HYGEFR----- 819
 283 LALPSDPOLHOKNDECAVCDGGLICCDGCPRAFFHACLSPPLREIPSGTWKSCSCLQ 342
 820 --KDDP-----NEDMCVAVONGGELLCCCKPKVFHLSCHVPTLTNPSGEMICFCMD 872
 343 ATVOEVOGRAEP 355
 873 LSKREVEYDCDAP 885

RESULT 7
 ID CHDM DROME STANDARD; PRT; 1982 AA.
 AC 097159; Q9VW50;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromodomain helicase-DNA-binding protein M1-2 homolog (M1-2).
 GN M1-2 OR CG8103.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NB1_Taxid=7227;
 [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS OF GLY-737.
 RX MEDLINE=99055400; PubMed=9836641;
 RA Kahle J., Beuchle D., Treuhelt S., Christen B., Kennison J.A.,
 RT Blenz M., Muller J.;
 RT "M1-2, a hunchback-interacting protein that functions in Polycomb
 repression";
 RL Science 282:1897-1900(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolishevsky S.,
 RA Borokova D., Buchan M.A., Butler H., Cadieu E., Center A.,
 RA Cherry J.M., Chew J., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Dugan-Rocha S., Dunn P.,
 RA Dudbin K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA Fostler K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Gaboriaud A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC - FUNCTION: VITAL ROLE IN DEVELOPMENT. PROTEIN BINDS TO A PORTION OF
 CC HUNCHBACK (HB) PROTEIN THAT IS CRITICAL FOR REPRESSION OF BITHORAX
 CC COMPLEX (BXC) GENES. MAY ALSO FUNCTION IN POLYCOMB GROUP (PCG)
 CC REPRESSION OF HOX GENES.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - SIMILARITY: BELONGS TO THE SNF2/RAD50 FAMILY.
 CC - SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

```

-! CAUTION: Ref.2 sequence differs from that shown due to erroneous
  gene model prediction.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb.sib.ch).
-----
EMBL: AF119716; AAD17276.1; -.
EMBL: AE003515; AAP49099.1; ALT_SEQ.
FlyBase: FBgn0013591; Mi-2.
InterPro: IPR000953; Chromo.
InterPro: IPR001410; DEAD.
InterPro: IPR002464; DEAD_box.
InterPro: IPR001650; Helicase_C.
InterPro: IPR000330; SNEF2_N.
InterPro: IPR001965; ZnF_PHD.
InterPro: IPR001841; ZnF_Ting.
Pfam: PF00176; SNEF2_N.1.
Pfam: PF00271; helicase_C.1.
Pfam: PF00385; chromo.1.
Pfam: PF00628; PHD.2.
SMART: SM00298; CHROMO.2.
SMART: SM00487; DEXDC.1.
SMART: SM00490; HELIC_C.1.
SMART: SM00249; PHD.2.
SMART: SM00184; RING.2.
PROSITE: PS00598; CHROMO.1; FALSE_NEG.
PROSITE: PS50013; CHROMO.2; 2.
PROSITE: PS00690; DEAD_ATP_HELICASE.1.
PROSITE: PS01359; ZF_PHD.1; 2.
PROSITE: PS50016; ZF_PHD.2; 2.
DNA-binding: ATP-binding; Helicase; Nuclear protein; Repeat;
KW Transcription regulation; Repressor; Zinc-finger.
FT ZN_RING 377 424
FT ZN_RING 437 484
FT ZN_RING 488 566
FT DOMAIN 612 673
FT NP_BIND 755 762
FT SITE 875 878
FT DOMAIN 13 16
FT DOMAIN 70 76
FT DOMAIN 239 248
FT DOMAIN 1279 1287
FT DOMAIN 1672 1677
FT MUTAGEN 737 737
FT CONFLICT 101 101
SQ
SEQUENCE 1982 AA; 224199 MW; ED8E256D1AD0ACAF CRC64;

Query Match 7.8%; Score 227; DB 1; Length 1982;
Best Local Similarity 20.2%; Pred. No. 2.7e-05;
Matches 77; Conservative 39; Mismatches 93; Indels 172; Gaps 11;

QY 110 RKGKRPAYPKALVPPRLPT-----RKASEEARAAPALPTPTGASPSQLK 159
   :|||:| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 260 KKGRRPSG-----KVPLTKIKLGGKKRRDSSDEEDAS-----CASRDSPL 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 160 AKPPKPESSAEQORLPLGKCIOTMSASVQRAVAMSSGDPGARCAVEGILLIOVPESGG 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 EMLQSDSDSADEKAEVAS-----SKADNSAPAA----- 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 220 SKKCIQVGGEFYIPRSKEDSGSGKNKARSSSGPKPLVRAKKQAGAPGGGEARLG----Q 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 -----QDDSG-----AVYRK-----AKTIGNKKFKK 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 276 QGSVPALPLPSDPLTHQKNEDBCAVCRDGGELLICDCDGRFAFHLACLSPLEIPSGTW 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 KKKLTKTKNPEEGEDGHNHODYCEVCGQGGEIILCDITCPRAVHLYACLEPDEDPREGKW 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 336 KSSSCIQATVQEQPRAEPRPQPPVETPLPGLRSAGEEYRGPPGEPGLAGMDTLVYK 395

```

Db	417	SCPHC-----EADGGAEEDEDDHQE-----	439
Qy	396	HLPPPSAAPLPGLDSSALHPLLGVPGGQONLAPGARGVCGDGYLRCTHCAAFHW	455
Db	440	-----	CRVCKDGDGELLCDSDSPSAVHT 461
Qy	456	RCHPAPAGTSRPGTGRCRSCS 476	
Db	462	FCLNPLPDITPDGDMRCPRCS 482	
RESULT 8			
TF1A_MOUSE			
ID	TF1A_MOUSE	STANDARD:	PRT: 1051 AA.
AC	Q64127; Q64126;		
DT	15-JUL-1999 (Rel. 38, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Transcription intermediary factor 1-alpha (TIF1-alpha).		
GN	TIF1A OR TIF1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
FN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Carcinoma;		
RX	MEDLINE=95262642; PubMed=7744009;		
RA	Le Douarin B., Zechel C., Garnier J.-M., Lutz Y., Tora L., Pierrat B.,		
RA	Heery D., Gronemeyer H., Chambon P., Losson R.;		
RT	"The N-terminal part of TIF1, a putative mediator of the ligand-		
RT	dependent activation function (AF-2) of nuclear receptors, is fused to		
RT	B rat 1n the oncogenic protein T18.";		
RL	EMBO J. 14:2020-2033(1995).		
CC	-1- FUNCTION: INTERACTS SELECTIVELY IN VITRO WITH THE AF2-ACTIVATING		
CC	DOMAIN OF THE ESTROGEN RECEPTORS. ASSOCIATION WITH DNA-BOUND		
CC	ESTROGEN RECEPTORS REQUIRES THE PRESENCE OF ESTRADIOL (BY		
CC	STIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: Nuclear.		
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A		
CC	SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	-1- TISSUE SPECIFICITY: DETECTED IN ALL ADULT TISSUES, WITH THE		
CC	HIGHEST EXPRESSION LEVEL IN TESTIS.		
CC	-1- DISEASE: PARTICIPATES IN A CHROMOSOMAL TRANSLOCATION THAT PRODUCES		
CC	A TIF1-BRAF (T18) ONCOGENE ORIGINALLY ISOLATED FROM A FURFURAL-		
CC	INDUCED HEPATOMA.		
CC	-1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.		
CC	-1- SIMILARITY: CONTAINS 2 B BOX-TYPE ZINC FINGERS.		
CC	-1- SIMILARITY: CONTAINS 1 BROMODOMAIN.		
CC	-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; S78221; AAB34290.1; -		
DR	EMBL; S78219; AAB34289.1; -		
DR	HSSP; P29590; IBOR.		
DR	TRANSFAC; T02143; -		
DR	TRANSFAC; T02215; -		
DR	MGD; MGI:109275; Tif1a.		
DR	InterPro; IPR0003649; Bbox_C.		
DR	InterPro; IPR001487; Bromodomain.		
DR	InterPro; IPR000315; Znf_Bbox.		
DR	InterPro; IPR001965; Znf_PHD.		
DR	InterPro; IPR001841; Znf_Ting.		
DR	Pfam; PF00097; zf-C3HC4; 1.		
DR	Pfam; PF00439; bromodomain; 1.		

DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00643; zf-B-box; 2.
 DR PRINTS: PRO1406; BROXZNRING.
 DR PRINTS: PRO0503; BROMODOMAIN.
 DR SMART: SM00502; BBOX; 1.
 DR SMART: SM00336; BBOX; 2.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS0014; BROMODOMAIN_2; 1.
 DR PROSITE: PS0119; ZF-BBOX; 2.
 DR PROSITE: PS01359; ZF-PHD; 1.
 DR PROSITE: PS0016; ZF-PHD; 2; 1.
 DR PROSITE: PS00518; ZF-RING; 1.
 DR PROSITE: PS00089; ZF-RING; 2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Bromodomain;
 KW Zinc-finger; Alternative splicing; Nuclear protein; Coiled coil;
 KW Repeat; Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 8 15
 FT ZN_FING 19 22 POLY-ALA.
 FT ZN_FING 52 77 RING-TYPE.
 FT ZN_FING 158 211 B BOX-TYPE 1.
 FT ZN_FING 218 259 B BOX-TYPE 2.
 FT DOMAIN 289 359 COILED COIL (POTENTIAL).
 FT DOMAIN 344 347 POLY-GLN.
 FT DOMAIN 583 587 POLY-SER.
 FT DOMAIN 755 780 NUCLEAR RECEPTOR BINDING SITE (NRBS).
 FT ZN_FING 827 874 PHD-TYPE.
 FT DOMAIN 892 908 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 933 988 BROMODOMAIN.
 FT SITE 332 333 BREAKPOINT FOR TRANSLLOCATION TO FORM
 FT SITE 333 333 TIPIA-BRAF ONCOGENE.
 FT VARSPLIC 477 510 MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 1051 AA; 116656 MW; 610584C1C685972 CRC64;

Query Match 7.5%; Score 217; DB 1; Length 1051;
 Best local similarity 27.2%; Pred. No. 5e-05;
 Matches 85; Conservative 46; Mismatches 122; Indels 60; Gaps 15;

QY 71 STALDWRVLFKDYLERGRLOPILDSFPRKVDLSQ-----PRKRRPPAVKALVPPP 126
 DB 606 SSPMIDLSAPVGSYN-----LPSLP-DIDCSSTIMLDNARKDTGVDAH--Qp 651
 QY 127 RLPTKRKASPEARAAAPALPLR-GTASPGQLAKPP-KKPESAEQRLPLGNGIOTM 184
 DB 652 RPPSNKTVQ-----SNSSVSPSPGLAGPVMTSVHPPIKNSPVSSVSGSSG----- 699
 QY 185 SASVQRAVAMSSGDVPGARGAVEGILIQVEESGSKK-----CIOVGGEFTYTSKFEDS 239
 DB 700 -SSSKPAGADSTHKVPV--MLEPIRIKQ--ENSGPPENDFPVYIVKQESDESRPONT 754
 QY 240 GSKK-----NKARSSSGPKPLVR--AKGAAGAPG-----GGEARLGQGSVPAP 282
 DB 755 NTPRSLITSLLLNMSQSSASEETVLRSDADPTGDLGHLQENSSNCKSE-WDASQKSP 813
 QY 283 LALPSPOLHOKEDECAYCRDGGELICCGCRAPHLAQLSPRLKIPSTGWRCCSCLO 342
 DB 814 VAVGETRKEDDEDKCAVQNGGELLCCCKCPVHLTCHVPLTTFPGSEWICTFCRD 873
 QY 343 ATVOEVQRAPEP 355
 DB 874 LSKPEVDYDCVP 886

RESULT 9
 CHD3-CAEEL STANDARD: PRT: 1787 AA.
 AC Q22516; Q18794;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromodomain helicase-DNA-binding protein 3 homolog (CHD-3).

GN CHD-3 OR T14G8.1.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20530483; PubMed=11076750;
 RA von Zellewsky T., Palladino F., Brunschwig K., Tobler H., Hajnal A.,
 RA Mueller F.,
 RT "The C. elegans Mi-2 chromatin-remodelling proteins function in vulval
 cell fate determination."
 RL Development 127:5277-5284(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Matthews P., McMurray A.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: Chromatin-remodelling protein that function in vulval
 cell fate determination.
 CC - SUBCELLULAR LOCATION: Nuclear (potential).
 CC - SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC - SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC - SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF308444; AAC29837.1; -
 DR EMBL: 267884; CAA1810.1; -
 DR EMBL: 267881; CAA91810.1; JOINED.
 DR EMBL: 267881; CAA91798.1; -
 DR EMBL: 267884; CAA91798.1; JOINED.
 DR Wormpep: T14G8.1; CE03657.
 DR Interpro: IPR000953; Chromo.
 DR Interpro: IPR001410; DEAD.
 DR Interpro: IPR002464; DEAH_box.
 DR Interpro: IPR001650; Helicase_C.
 DR Interpro: IPR000330; SNF2_N.
 DR Interpro: IPR001965; znf-PHD.
 DR Interpro: IPR001841; znf_ring.
 DR Pfam: PF00176; SNF2_N; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00385; chromo; 1.
 DR Pfam: PF00628; PHD; 2.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR PROSITE: PS00598; CHROMO_1; FALSE_NEG.
 DR PROSITE: PS00513; CHROMO_2; 1.
 DR PROSITE: PS00690; DEAH-ATP-HELICASE; 1.
 DR PROSITE: PS01359; ZF-PHD; 1; 2.
 DR PROSITE: PS0016; ZF-PHD; 2; 2.
 KW Chromatin regulator; Nuclear protein; Repeat; Helicase; DNA-binding;
 KW ATP-binding; Zinc-finger.
 FT DOMAIN 59 62 POLY-LYS.
 FT ZN_FING 265 312 PHD-TYPE 1.
 FT ZN_FING 328 375 PHD-TYPE 2.
 FT DOMAIN 373 476 CHROMO 1.
 FT DOMAIN 501 583 CHROMO 2.
 FT DOMAIN 1287 1291 POLY-ARG.
 FT NP_BIND 641 648 ATP (POTENTIAL).
 FT SITE 763 766 DEAH BOX.
 FT SITE 763 766
 FT SEQUENCE 1787 AA; 205254 MW; 1EFCE1FECE59740 CRC64;

Query Match 7.1%; Score 206; DB 1; Length 1787;

QY 227 GGE-----FYTSPKFEEDSGSGKNKARS-----SSGPKPLVRAKGAQGAAP 266
 DB 809 SPESLTPPLSTNLHESLDLALASLENNHYKTEPADMNESSCKOSGLSLVNGKSPINSLM 868
 QY 267 GGGEARLGGCGSVAPLALPSPDLHOKNEDECAVCRDGGELICDCCPRAFHACTLSP 326
 DB 869 -HRSARIGGDGN-----NKDDP-----NEDMCATVCNGGDLCCCEKCPKFKHLTCHVPT 917
 QY 327 LREIPSGTWRCSSCLQATVQEVQ 349
 DB 918 LLSFSGDWICTCFCDKRCKEVE 940
 RESULT 12
 YAZ7_SCHPO STANDARD; PRT; 607 AA.
 AC 009698;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C2E7.07c in chromosome I.
 GN SPAC2E7.07c.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
 Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 Borzym K., Zimmermann W., Wedler H., Reinhardt R., Pohl T.M.,
 Eger P., Zimmermann W., Wedler H., Reinhardt R., Pohl T.M.,
 Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Motier S.,
 Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 Lucas R., Rochet M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Paulsen I., Potashkin J.,
 Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -1- SIMILARITY: TO S.POMBE SPAC16C9.05 AND YEAST YMR075W.
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z50142; CAA90494.1;
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 2.
 DR SMART: SM00249; PHD; 2.

DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS50016; ZF_PHD_2; 2.
 KM Hypothetical protein; Repeat; Zinc-finger.
 FT ZN_FING 263 312 PHD-TYPE 1.
 FT ZN_FING 406 459 PHD-TYPE 2.
 SQ SEQUENCE 607 AA; 68776 MW; 43CB6520BA471AC CRC64;
 Query Match 6.4%; Score 186.5; DB 1; Length 607;
 Best Local Similarity 22.7%; Pred. No. 0.0014;
 Matches 83; Conservative 44; Mismatches 13; Indels 107; Gaps 18;
 QY 151 TASPGLAKAKPKPPSSAEQRLPLNGIOTMSAVORAVAMSSGDVPGARAVEGIL 210
 DB 169 TAKDLSDISSSMKKANNSK-----PLPSGLTFKANI-----PVPTSEVVTENNV 215
 QY 211 IQQYFESGSKKCIQVGEFTTPSKFEEDSGSGKNKARS--GKPLVRAKGAQGAAPGGE 270
 DB 216 TRNVTYVSNOK---HIGNE---SENFNDM---EGRAEDISSNEL----- 251
 QY 271 ARLGQCGSVAPLALPSPDLHOKNEDECAVCRDGGELICDCCPRAFHACTLSP 329
 DB 252 -----FTPEYTP-----YRINDYCSACHGPNFLCCTCNSHFHTCIDPIEEK 297
 QY 330 -IPSGTWRCSSCLQATV-----QEVQPRAEPR-----POEPPVETPLDP 368
 DB 298 NLPDDAWYCNCKRHSLYNELEDEQELSENVKEEGTVDVMMQLCTYIDSHNPIDFLPH 357
 QY 369 GL-----RSAGEEVGPRGPEPLAGMDTLYVKHLRPA-----PSAAPPLDLSALHPLLC 419
 DB 358 SISEFFRGVSGVGEYIE-----TDVLRKLKSSRSNGERDPL-LKSKSGTPIIC 409
 QY 420 VGEPEGQNLAPGARGVCG-DGTDLVLRCTHCAAHWRHCFPAGTSRPCTGLR---CRSC 475
 DB 410 F-----RCHSALVQSITLADCYGNSYHNPDLNPLATLP-SNLKMKMCPHN 456
 QY 476 SGDVTP 481
 DB 457 SDHYTP 462
 RESULT 13
 YAZ5_SCHPO STANDARD; PRT; 404 AA.
 AC 009819;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C16C9.05 in chromosome I.
 GN SPAC16C9.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
 Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard J., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Gurtelt L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipkovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schistosoma mansoni pombae";
 RL Nature 415:871-880(2002).
 CC -1-SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 254366; CAA01193.1;
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD_1.
 DR SMART: SM00249; PHD_1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 DR KW Hypothetical protein; Zinc-finger.
 FT ZN_FING 117 166 PHD-TYPE.
 FT SEQUENCE 404 AA; 44984 MW; CB9412E792C1C5D4 CRC64;

Query Match

Best Local Similarity 6.2%; Score 179.5; DB 1; Length 404;
 Matches 67; Conservative 27; Mismatches 94; Indels 75; Gaps 12;

QY 234 KNEDECAVCRDGGELICCGCPAPHLACSLPRL--RETPSGTWKSSCCQATVQEVQPR 351
 DB 115 RNVDICSNACGGGRGIFCCGCPSCFHLSCLEPLTPENIPREGSMFCVTC--SIKSHHP 170
 QY 352 AEEPRQEP-----PVEPTLPRLPGILRSAGEEV--RGPG--EPLAGMDTTL 392
 DB 171 -----PKHPLSLTWSQLYDWIDQNSQYRLPDDLVHYHGISGDMTGAKYKETEEMDT-- 223
 QY 393 YKKHLPAAPSAAPRLGGLSSALHPLLCVGRPEGOQLNAPARCGVCGDGT-----VLRCSTH 448
 DB 224 --DEFSALPTGSSITNL-----AYCGYCKSPMGACVWYGCQL 259
 QY 449 CAAAFHWKCHPAGTSRPGTGLRCSCGSDYTPAPVEGVLAAPARLAPGAKDPTASH-- 507
 DB 260 CDTEYHNKCKEHA-----KKCSHD--SIKKGRVPEKMAVIRTPPLVLDTSMT 306
 QY 508 -EPALHNDLESLSSEHTFDGIL 529
 DB 307 LNPVKVMSIGWQFLMGEPSSDEL 329

RESULT 14

YHLL_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DE 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical BHLF1 protein.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OK NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,

RA Tuftnell P.S., Barrell B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome";
 RL Nature 310:207-211(1994).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: A01555; NOT_ANNOTATED_CDS.
 DR PIR: A03742; Q0BE3.
 KW Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 149 648 4 x 125 AA TANDDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 FT SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Query Match

Best Local Similarity 5.8%; Score 168.5; DB 1; Length 660;
 Matches 120; Conservative 26; Mismatches 192; Indels 175; Gaps 25;

QY 109 PKRGRK-----PAVRLATVPRLPTKRRASEEAAARAALPLR-----GTASPGS 156
 DB 172 PGAGRGPRSGPTGGRRAARPA-----PQTPA--APGCGGAAYVSGATPHERGSGPADPPA 225
 QY 157 QLKAKP-----PKPESSAEQRLPLGNGIOTMSASAVQAVAMSSGDVPCAR----- 203
 DB 226 AARLPREPEREPLRQDLAAQRCRPAQRPPTTSGAAQRTNRRRRCGRRAARNDGCRTRR 285
 QY 204 ---GAVEGILLQVVESSGSKKCIQVGEFTYPSKFEDSGSKNKARSSGKRLVRAKG 260
 DB 286 RRSQAQR-----HPP-----PGAQ--RPSGPTGGRP--AARG 315
 QY 261 AAG--AAPGGEARLGGQGSYAPRL--ALPSPDLQKNEDCAVCRDGGELICDGCPR 316
 DB 316 APTRAAPRGFGGAAYVSGATPHERGSGPADPPAALPLPREQERPLRQDLAAQRCRPA 375
 QY 317 AFHILACSLPRL-----ETPSG-----TWRCSSCQATVQEVQPRAEPR 355
 DB 376 G-----PPPTTSGAAQRTNRRRRCGRRAARNDGCRTRRRAQAQ----- 416
 QY 356 RPOEPVETPLPRLRSAGEVRRGPRGLAGMDTTLVYKKLPAAPSAAPRLGGLDSSALH 415
 DB 417 RGNPRPGAGRPSGPTGGRRAARCAPTPRA-----PGGGGAAYV--SGATPH 463
 QY 416 PLLCVG-----PEGO-----QNLAPGARC-----GYCGDGT----- 442
 DB 464 PERGSGPADPPAARLPRERQERPLRQDLAAQRCRPAQRPPTTSGAAQRTNRRRRCGR 523
 QY 443 VLRCSTCAAFHWKCHPAGTSRPGTGLRCSCG-----DVTAPVEGVLAAP 491
 DB 524 SARNPCCPTTWRRRSGAQRCNRPFGAGQRPSTGGRRAARCAPTPRAAPRGGAAYVPS 583
 QY 492 PARLAP-----GPAKDDTAS-----HEPALHRD 514
 DB 584 GATPIPERGSGPADPPAARLPRERQERPLRQD 616

RESULT 15

YMW5_YEAST STANDARD; PRT; 684 AA.
 AC Q04779;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 78.8 kDa protein in ABF2-CHL12 intergenic region.
 GN YMR075W OR YW9916.14.
 OS Saccharomyces cerevisiae (Baker's yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO S.POMBE SPAC16C9.05 AND SPAC2F7.07C.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; 248952; CA88800.1; -.
DR SGD; S0004680; YMR075W.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Hypothetical protein; Zinc-finger.
FT ZN_FING 260 309 PHD-TYPE.
FT DOMAIN 15 30 POLY-SER.
SQ SEQUENCE 684 AA: 78836 MW: CRC282DC9A782E5C CRC64;

Query Match 5.5%; Score 161; DB 1; Length 684;
Best Local Similarity 21.7%; Pred. No. 0.038;
Matches 48; Conservative 34; Mismatches 85; Indels 54; Gaps 8;

QY 294 KNEDECAVCRDGGELICDGGPRAFHACLSPRL--REIPSGTWRCSSC-----LQA 343
DB 258 ENEDFCSACNOSGSEFLCCDTCPSHFHCLDPPIDPNNILPKGDWHCNECKFKIFINNSMA 317
QY 344 TVQEVQPRAEPR-----POEPVETPLPPGLRSAGEVR-GPGE----- 383
DB 318 TLKKIESNFIKONNNVAKIFAKLLFNIDSHNPKQFOLPNYIKETFPRAVKTSRGQYSDEND 377
QY 384 --PLAG--MDTLYYKHLPAFSAAPLPGLDSSALHPLLCVGPPEGQONLAPGARCGVCGD 439
DB 378 KIPLFTDRQLFNTSYGQSITKLDLSYNDPTNHDINSNGKFLICY-----KCNQTRL 425
QY 440 GT-----DVLRCTHCAAFHWRCHPRPAGTSRGTGLRC 472
DB 426 GSWSHPENSRLLIMTCDYCQTFPHLDYCPRASFKNLGSKWK 466

Search completed: March 13, 2003, 17:53:56
Job time : 20 secs

